

Durham Research Online

Deposited in DRO:

07 December 2021

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Hadzhiyska, Boryana and Liu, Sonya and Somerville, Rachel S and Gabrielpillai, Austen and Bose, Sowndak and Eisenstein, Daniel and Hernquist, Lars (2021) 'Galaxy assembly bias and large-scale distribution: a comparison between IllustrisTNG and a semi-analytic model.', *Monthly notices of the Royal Astronomical Society*, 508 (1). pp. 698-718.

Further information on publisher's website:

<https://doi.org/10.1093/mnras/stab2564>

Publisher's copyright statement:

This article has been accepted for publication in *Monthly Notices of the Royal Astronomical Society* ©: 2021, The Authors. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Galaxy assembly bias and large-scale distribution: a comparison between IllustrisTNG and a semi-analytic model

Boryana Hadzhiyska¹,¹★ Sonya Liu,² Rachel S. Somerville,³ Austen Gabrielpillai,³ Sownak Bose¹,¹ Daniel Eisenstein¹ and Lars Hernquist¹

¹Center for Astrophysics, Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA

²American Heritage School, 736 N 1100 E, American Fork, UT 84003, USA

³Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA

Accepted 2021 September 2. Received 2021 August 23; in original form 2021 July 7

ABSTRACT

In this work, we compare large scale structure observables for stellar mass selected samples at $z = 0$, as predicted by two galaxy models, the hydrodynamical simulation IllustrisTNG and the Santa-Cruz semi-analytic model (SC-SAM). Although both models have been independently calibrated to match observations, rather than each other, we find good agreement between the two models for two-point clustering and galaxy assembly bias signatures. The models also show a qualitatively similar response of occupancy and clustering to secondary halo parameters other than mass, such as formation history and concentration, although with some quantitative differences. Thus, our results demonstrate that the galaxy–halo relationships in SC-SAM and TNG are quite similar to first order. However, we also find areas in which the models differ. For example, we note a strong correlation between halo gas content and environment in TNG, which is lacking in the SC-SAM, as well as differences in the occupancy predictions for low-mass haloes. Moreover, we show that higher order statistics, such as cumulants of the density field, help us to accurately describe the galaxy distribution and discriminate between models that show degenerate behaviour for two-point statistics. Our results suggest that SAMs are a promising cost-effective and intuitive method for generating mock catalogues for next generation cosmological surveys.

Key words: galaxies: formation – galaxies: haloes – large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

According to the concordance model of big bang cosmology, our Universe is made up of a network of filaments, shaped by gravity and populated with dark matter haloes. These dark matter haloes correspond to overdense regions, which have evolved by gravitational instability and interactions with other haloes. In this framework, galaxy formation takes place within such haloes, as baryonic matter sinks to the centre of their potential wells and cold gas condenses to form galaxies (White & Rees 1978). Since we cannot observe dark matter directly, understanding the galaxy–halo connection by studying the distribution of galaxies would enable us to place stringent constraints on cosmological parameters.

The formation and evolution of dark matter haloes can be modelled with a high degree of accuracy by cosmological (N -body) simulations. Such calculations incorporate various assumptions of the standard paradigm and adopt particular values for the cosmological parameters, which are often determined by observations. They have the benefit of being relatively computationally inexpensive and, thus, can encompass very large volumes ($\sim 1 \text{ Gpc } h^{-1}$); a necessity to perform high-fidelity forecasts and interpretation of current and future galaxy surveys. However, a considerable shortcoming of these simulations is that they lack prescriptions for determining the detailed formation and evolutionary history of the observable components of

galaxies. To mimic the galactic component in such large volumes, cosmologists often adopt a posteriori models of varying complexity to populate the dark matter haloes with galaxies.

The simplest of these models involve phenomenological approaches, such as halo occupation distribution (HOD) models, which provide an empirical relation between halo mass and the number of galaxies in a host halo (Benson et al. 2000; Peacock & Smith 2000; Scoccimarro et al. 2001; Yang et al. 2004; Wechsler & Tinker 2018). HOD modelling is one of the most popular ways to ‘paint’ galaxies on to N -body simulations. The process of generating multiple mock catalogues in large volumes is very efficient and can thus satisfy the demanding requirements of next-generation experiments. Unfortunately, the simplest versions of such empirical approaches have been shown to make predictions that have significant discrepancies with observations and more detailed galaxy formation models (Croton, Gao & White 2007; Paranjape et al. 2015; Beltz-Mohrmann, Berlind & Szeziwciw 2020). For example, several recent works have argued that including secondary HOD parameters, such as the local halo environment and concentration, yields better agreement with hydrodynamical simulations as compared to the standard prescription (Hadzhiyska et al. 2020; Xu & Zheng 2020; Hadzhiyska et al. 2021a). Related approaches include Sub-Halo Abundance Matching models (SHAMs) and semi-empirical models, which derive mappings between galaxy and DM halo properties such that specific observational quantities are matched (Behroozi, Wechsler & Conroy 2013b; Moster, Naab & White 2013, 2018; Tacchella, Trenti & Carollo 2013; Tacchella

* E-mail: boryanah@alumni.princeton.edu

et al. 2018; Behroozi et al. 2019). These methods can also be effective for efficient construction of realistic galaxy mock catalogues. A common feature of HOD, SHAM, and semi-empirical models is that they derive empirical mappings between galaxies and DM haloes, without attempting to model *ab initio* physical processes.

On the other hand, numerical hydrodynamical simulations provide an *ab initio* approach for gaining insight into the evolution and distribution of galaxies (e.g. Vogelsberger et al. 2014a; Schaye et al. 2015), incorporating a broad suite of physical processes, including gravity, hydrodynamics, and thermodynamics. These simulations also incorporate key baryonic processes such as star formation, stellar winds, chemical enrichment, and black hole (BH) growth and feedback. However, due to the limited dynamic range of the simulations, such processes must be treated with ‘sub-grid’ recipes which describe physics that occurs below the explicit resolution of the numerical scheme. These sub-grid processes are typically phenomenological, and contain adjustable parameters which are typically calibrated to reproduce a set of observations. Having recently reached sizes of several hundred megaparsecs on a side (Chaves-Montero et al. 2016; Springel et al. 2018), hydro simulations now allow us to test the accuracy of galaxy population models by comparing various statistical properties of the galaxy field side by side (Bose et al. 2019; Hadzhiyska et al. 2020; Contreras, Angulo & Zennaro 2021). Such analyses have been performed extensively for galaxy populations above a certain stellar mass cut (as stellar-mass cuts are a good proxy for luminosity cuts), and a few studies have focused on other galaxy populations targeted by future surveys, such as star-forming emission-line galaxies (ELGs) (e.g. Zheng et al. 2005; Guo et al. 2016b; Favole et al. 2017; Cochrane & Best 2018; Gonzalez-Perez et al. 2018; Alam et al. 2019; Hadzhiyska et al. 2021b). While hydro simulations can be used to inform us about the galaxy–halo connection on intermediate and quasi-linear scales, they are computationally expensive and cannot (at present) be run in the large volumes needed for future cosmological surveys. Moreover, the predictions for galaxy properties are sensitive to the details of the sub-grid prescriptions, and it is too expensive to explore many variations of these in large volumes.

A particularly promising and well-explored route for physics-based yet more computationally efficient modelling of the galaxy–halo connection is through semi-analytic models (SAMs) of the galaxy formation. SAMs are built using halo merger trees extracted from an N -body simulation and employ various physical prescriptions to predict the evolution of baryons and their flow between different ‘reservoirs’, such as the intergalactic medium (IGM), circumgalactic medium (CGM), interstellar medium (ISM), and galactic stellar bulge, disc, and halo (for a review, see Somerville & Davé 2015). A caveat of these models is that the evolution of baryonic matter is not obtained by directly solving the equations of gravity, hydrodynamics, thermodynamics, etc., but rather approximated through analytic calculations. Similarly to cosmological hydro simulations, key physical processes such as star formation, stellar feedback, and BH feedback are treated with parametrized phenomenological recipes, which are calibrated to match a subset of observations.

Understanding the connection between galaxies and their host dark matter haloes will help us obtain more realistic mock catalogues, which are crucial for the analysis of future observational surveys. One of the main limitations of ubiquitously adopted galaxy–halo approaches, such as the HOD model, is the amount of ‘galaxy assembly bias’ they predict. Galaxy assembly bias is the relative increase (or decrease) of the large-scale clustering of a galaxy sample resulting from dependencies of the galaxy–halo connection on halo

properties beyond mass, such as assembly history and environment (e.g. Croton et al. 2007). Thus, the galaxy assembly bias signal is the consequence of the effects of both halo assembly bias and occupation variations. The former is the dependence of halo clustering on properties other than mass (e.g. Gao, Springel & White 2005). The latter is the dependence of the properties of the galaxies within a halo on halo properties in addition to mass (see Artale et al. 2018; Zehavi et al. 2018). Some of the properties thought to contribute to the galaxy assembly bias signal are formation time, concentration, spin, and environment. Various authors have compared observations and realistic galaxy models with mock catalogues generated with and without taking assembly bias into consideration in an effort to study the strength of that effect (Contreras et al. 2019).

In this paper, we explore the large-scale galaxy distributions predicted by the numerical hydrodynamic simulation IllustrisTNG (Weinberger et al. 2017; Pillepich et al. 2018a) and the Santa Cruz semi-analytic model (Somerville & Primack 1999; Somerville et al. 2008; Somerville, Popping & Trager 2015; Gabrielpillai et al., in preparation; Somerville et al. 2021). Multiple analyses in the past have studied the statistical properties of galaxy populations predicted by different sets of hydrodynamical simulations and semi-analytic models. For example, Guo et al. (2016b) compared the Eagle hydrodynamical simulation (Schaye et al. 2015) with the L-GALAXIES SAM (Guo et al. 2013; Henriques et al. 2015) and Galform SAM (Gonzalez-Perez et al. 2014). Wang et al. (2018) compared both Eagle and L-GALAXIES with observations of quenched fractions. Popping et al. (2019) compared cold gas contents in the Santa Cruz SAM and the IllustrisTNG hydrodynamic simulations with observations. Pandya et al. (2020) compared the Santa Cruz SAM with the FIRE-2 hydrodynamic simulations (Hopkins et al. 2018). Most relevant to this study are the works by Renneby et al. (2020), who compared lensing and clustering probes in IllustrisTNG and the L-GALAXIES SAM with observations, and Contreras et al. (2021), who studied large-scale structure in IllustrisTNG and the SAGE semi-analytic model (Croton et al. 2016). We summarize their findings and compare them with this study in Section 4.

This paper is structured as follows. In Section 2, we summarize the semi-analytic model adopted in this work, SC-SAM, and briefly describe the IllustrisTNG suite of simulations. Moreover, we define various secondary halo parameters considered in our study of the assembly bias signal as well as the galaxy samples for which we compute summary statistics. In Section 3, we compare the large-scale properties of the TNG and SAM samples. In particular, we explore the halo occupation distributions, the clustering and the assembly bias of our galaxy samples, and the variation caused by secondary parameters. We also examine the spatial distribution of the galaxies and the amount of gas in the host halo conditioning on its environment. Finally, we analyse the differences in the clustering properties of galaxies when using alternative statistical probes, such as galaxy–galaxy lensing, galaxy–void correlation functions and the cumulants of the smoothed galaxy density. We discuss the interpretation and implications of our results in Section 4, and our conclusions are summarized in Section 5.

2 METHODOLOGY

In this section, we describe the Santa-Cruz semi-analytic model and the hydrodynamical simulation, IllustrisTNG, used in this work. The main focus of this study is on mass-selected galaxy samples at $z = 0$, as predicted by both models. Other types of galaxy tracers at early redshifts will be explored in future work.

2.1 Santa Cruz semi-analytic model

The version of the Santa Cruz SAM adopted in this study is the same as Gabrielpillai et al. (in preparation) and is similar to SAMs published in Somerville et al. (2015) and Somerville et al. (2021). Interested readers are pointed to these references, as well as Somerville et al. (2008) for more information.

The semi-analytic model is run on merger trees created using the ROCKSTAR halo-finding algorithm (Behroozi, Wechsler & Wu 2013a). ROCKSTAR is a temporal, phase-space finder; due to its use of both the phase space distribution of particles and temporal evolution, it is considered highly accurate in determining particle-halo membership. Having information about the relative motion of haloes makes the process of finding tidal remnants and determining halo boundaries substantially more effective, while having temporal information helps maximize the consistency of halo properties across time, rather than just within a single snapshot. The parameters chosen to construct the merger trees are the default ROCKSTAR parameters.¹ The ROCKSTAR merger trees provide information on the growth of collapsed dark matter haloes over time through accretion of ‘diffuse’ material and mergers with other haloes.

The rate of gas flow into a halo is given by the product of the universal baryon fraction and the growth rate of the dark matter halo, i.e. $f_b M_h$. This relation holds in the absence of feedback processes and takes into account the suppression of gas inflow due to photoionization through the expression for the collapsed baryon fraction:

$$f_{b,\text{coll}} = \frac{f_b}{[1 + 0.26 M_F(z)/M_{\text{vir}}]^3}, \quad (1)$$

where M_{vir} is the halo virial mass and M_F is the filtering mass, which depends on the reionization history of the Universe. The accreted gas forms the so-called ‘hot halo’ and is assumed to be isothermal and at the virial temperature. The ‘cooling radius’ is computed using standard radiative cooling functions from Sutherland & Dopita (1993). When the cooling radius is smaller than the virial radius, r_{vir} , a standard cooling flow model is applied. The cooling time is calculated via equation (1) in Somerville et al. (2008)

$$t_{\text{cool}} = \frac{(3/2) \mu m_p kT}{\rho_g(r) \Lambda(T, Z_h)}, \quad (2)$$

where μm_p is the mean molecular mass, $kT \propto V_{\text{vir}}^2$ is the virial temperature, $\rho_g = m_{\text{hot}}/(4\pi r_{\text{vir}}^2)$ is the gas density profile, and $\Lambda(T, Z_h)$ is the temperature and metallicity-dependent cooling function. As the cooling radius moves outward, the gas contained within it is assumed to cool and accrete on to the disc. When the cooling radius is larger than r_{vir} , the cooling rate is given by the gas accretion rate into the halo.

The star formation rate of galaxies is computed by assuming a scaling relation between the molecular gas surface density and star formation rate surface density, as supported by observations. The gas profile is assumed to be an exponential disc, proportional to the stellar disc, r_{disc} , where r_{disc} is computed by assuming angular momentum conservation and utilizing the halo parameters for concentration c_{NFW} , spin λ , and the fraction of baryons in the disc f_{disc} . Stellar feedback can eject cold gas from the ISM, with a rate determined by the circular velocity of the halo at twice the Navarro–Frenk–White (NFW; Navarro, Frenk & White 1996) scale radius, V_{disc} , the star

formation rate \dot{m}_* , and adjustable parameters ($\epsilon_{\text{SN}} = 1.7$, $\alpha_{\text{th}} = 3$):

$$\dot{m}_{\text{eject}} = \epsilon_{\text{SN}} \left(\frac{200 \text{ km s}^{-1}}{V_{\text{disc}}} \right)^{\alpha_{\text{th}}} \dot{m}_*. \quad (3)$$

A fraction of this gas is ejected from the halo and stored in an ‘ejected’ reservoir, while the rest is deposited in the hot gas halo. Gas in the ejected reservoir is ‘re-accreted’ at rate dependent on the halo dynamical time, $t_{\text{dyn}} = r_{\text{vir}}/V_{\text{vir}}$, with a normalization that contains another adjustable parameter.

Each top-level halo in the merger tree (i.e. initial progenitor) is seeded with a black hole, which can accrete mass through cooling flows from the hot halo or by inflows of cold gas driven by mergers or internal gravitational instabilities. Cold gas can be ejected from the ISM of the galaxy through AGN-driven winds triggered by mergers or disc instabilities. Gas ejected by the AGN is assumed to be permanently ejected from the parent halo, and does not get recycled. The radiatively inefficient mode of growth (‘radio mode’) generates a heating term, which can partially or completely offset cooling and accretion from the hot halo. For more details on the physics recipes adopted in SC-SAM, see Somerville et al. (2008).

Galaxies residing at the centre of their host haloes are considered ‘centrals’, while those hosted by subhaloes, as defined by ROCKSTAR, are called ‘satellites’. Rather than relying on N -body simulations for information about the subhaloes, the Santa-Cruz SAM tracks the orbital decay and tidal destruction of satellite galaxies utilizing a semi-analytic prescription (see Somerville et al. 2008, for details).

The SC SAM is tuned to match a subset of observations at $z = 0$ only, which are: the stellar mass function (equivalent to matching the stellar mass versus halo mass relation), the stellar mass versus cold gas fraction, stellar mass versus metallicity relation, and the bulge mass versus black hole mass relation. Please see Yung et al. (2019) Appendix B and Gabrielpillai et al. (in preparation) for details of the observations used for calibration.

The SC-SAM is run on the dark-matter-only IllustrisTNG simulations, so an additional product of the SC-SAM runs presented in Gabrielpillai et al. (in preparation) is a catalogue of halo and subhalo bijective matches between IllustrisTNG and SC-SAM, which plays a crucial role in determining the dependence of large-scale clustering observables on the halo assembly bias of both models. We also utilize these matches to provide more realistic positions to the SC-SAM satellites, since as noted above, the SAM does not use the subhalo positions and velocities from the N -body simulation. By taking the subhalo positions for a given ROCKSTAR halo directly from its bijectively matched FoF group, the positions of satellites in the SC-SAM correspond to the true positions of subhaloes belonging to that halo, as identified in the TNG full-physics run. The effect of this exercise is two-fold: on one hand, it simplifies the comparison of the one-halo term in TNG to SC-SAM; on the other, it enables us to examine the weak lensing signal of the two models, as the SC-SAM satellite distribution is correlated with the dark-matter distribution.

2.2 IllustrisTNG

The Next Generation Illustris simulation (IllustrisTNG, TNG), which is run with the AREPO code (Springel 2010; Weinberger, Springel & Pakmor 2019), consists of 9 simulations: 3 box sizes (300, 100, and 50 Mpc on a side), each available at 3 different resolutions, 1–3, with 1 being the highest and 3 the lowest resolution (Marinacci et al. 2018; Naiman et al. 2018; Springel et al. 2018; Nelson et al. 2019; Pillepich et al. 2019). Compared with its predecessor, Illustris (Genel et al. 2014; Vogelsberger et al. 2014a,b), TNG provides improved

¹For more information, see <https://github.com/yt-project/rockstar/blob/master/README.md>.

agreement with observations by modifying its treatment of active galactic nuclei (AGNs) feedback, galactic winds, and magnetic fields (Weinberger et al. 2017; Pillepich et al. 2018a). In addition, various improvements of the flexibility and hydrodynamical convergence have been introduced in the code.

Several physical processes cannot be directly simulated in cosmological hydrodynamical simulations, due to resolution limitations. These are implemented using ‘sub-grid’ recipes, and include star formation, stellar feedback, black hole seeding, black hole growth, and black hole feedback. Briefly, star formation and pressurization of the Interstellar Medium (ISM) is modelled using the approach of Springel & Hernquist (2003), as in the original Illustris simulation. Star formation drives galactic winds by injecting kinetic energy into gas cells with a probability governed by a parametrized function that depends on the gas metallicity and the dark matter (DM) velocity dispersion σ_{DM} , measured with a weighted kernel over the nearest 64 DM particles. The velocity imparted on to the wind particles is also a parametrized function of σ_{DM} and redshift. Seed black holes of a fixed mass are deposited in haloes above a critical mass, and BH accretion is modelled using a standard Bondi–Hoyle model. When the BH accretion rate is below a critical Eddington rate and the BH mass is above a critical value, highly energetic BH-driven winds are launched, again by depositing kinetic energy into gas cells near the BH. Further details are discussed in Pillepich et al. (2018a) and Weinberger et al. (2017).

The parameters that characterize the sub-grid processes in TNG have been calibrated to approximately reproduce a set of observations. Although the philosophy is similar to that used to calibrate SAMs, the details of which observations are used for calibration, and the required precision of the calibration, can be different. The observations used for calibration of TNG are discussed in Pillepich et al. (2018a), and include the cosmic SFR density as a function of redshift, stellar mass function, BH mass versus stellar mass relation, hot gas fraction in galaxy clusters, and galaxy stellar mass versus radius relation. In addition, the distribution of galaxy optical colours in stellar mass bins was also used to motivate and calibrate the AGN feedback parameters as discussed in Nelson et al. (2018). We note that these choices overlap with, but differ from, those used to calibrate the SC SAM.

In this paper, we use the largest box ($L_{\text{box}} = 205 \text{ Mpc } h^{-1}$) at its highest resolution, TNG300-1, with dark matter particle mass of $3.98 \times 10^7 M_{\odot} h^{-1}$ and baryonic (gas cells and star particles) mass resolution of $7.44 \times 10^6 M_{\odot} h^{-1}$. TNG provides both the hydrodynamical (full-physics) and the dark matter only (N -body) simulation output, evolved from the same set of initial conditions. It is crucial to test any hypotheses regarding the galaxy–halo connection on the dark-matter-only counterpart, since mock catalogues used in cosmological galaxy surveys typically utilize cosmological N -body simulations, as they can be produced in sufficiently large volumes. Having a bijective match between the full-physics and N -body runs also gives us an opportunity to make halo-by-halo comparisons and thus study the effect of baryons on the properties of haloes.

Haloes (groups) in TNG are identified using the standard friends-of-friends (FoF; Davis et al. 1985; Barnes & Efstathiou 1987) algorithm, which depends on a single parameter, the linking length $b = 0.2$, and is run on the dark matter particles. The outer boundaries of the FoF haloes roughly correspond to an overdensity contour of ~ 180 times the background density using the percolation theory results of More et al. (2011). However, haloes identified by this method may appear as two or more clumps, linked by a small thread of particles, which leads to a miscalculation of their properties. The subhaloes, on the other hand, are found using the SUBFIND

algorithm (Springel et al. 2001), which detects substructure within the groups and defines locally overdense, self-bound particle groups.

2.3 Selection of the galaxy samples

In this work, we consider stellar-mass-selected galaxy samples at $z = 0$. Although such mass-selected samples do not necessarily satisfy the same set of magnitude and colour cuts that luminous red galaxies (LRGs) must satisfy to be targeted by modern galaxy surveys (e.g. DESI, SDSS, *Euclid*), they are believed to largely overlap with the LRGs selected in imaging surveys (Masaki et al. 2013). The stellar mass selection is performed at three different galaxy number densities, aimed to mimic current and near-future galaxy surveys, $n_{\text{gal}} = 0.0005, 0.001, \text{ and } 0.002 [\text{Mpc } h^{-1}]^{-3}$. The TNG subhalo stellar mass definition found to most closely correspond to the stellar mass of the galaxy bulge and disc reported in SC-SAM is ‘SubhaloMassInRadType’; i.e. the sum of masses of all ‘star’ particles within twice the stellar half-mass radius. The lowest stellar mass of the TNG sample is $\log M_* = 10.56$, while that of SC-SAM is $\log M_* = 10.81$, expressed in units of $M_{\odot} h^{-1}$. The TNG halo mass that we adopt is ‘Group_M_TopHat200’ ($\equiv M_{200}$), i.e. the total mass enclosed in a sphere whose mean density is Δ_c times the critical density of the Universe. Δ_c is derived from the solution of the collapse of a spherical top-hat perturbation. A well-known fitting function is provided by

$$\Delta_c(z) = 18\pi^2 + 82x - 39x^2, \quad (4)$$

where $x = \Omega_m(z) - 1$ and $\Omega_m(z)$ is the matter energy density at redshift z (Bryan & Norman 1998). The virial mass of ROCKSTAR haloes is defined using the same redshift scaling of Δ_c .

We construct the TNG galaxy sample from the most massive N galaxies, where $N = 6000, 12\,000, \text{ and } 24\,000$, corresponding to the three density thresholds above. This results in roughly a 3:1 split of centrals to satellites as defined by SUBFIND. Rather than utilizing the same indiscriminate selection in SC-SAM, we require that the satellite-central ratio be the same in SC-SAM when selecting the sample of N galaxies, ordered by stellar mass. Thus, we obtain two galaxy samples that have matching satellite fractions and galaxy densities. The constraint of a fixed satellite fraction is enforced due to a fundamental difference between the two models in the treatment of satellites. First, the merging, tidal stripping, and destruction of satellites in the SC-SAM is treated using a semi-analytic model. Secondly, once a halo becomes a subhalo (satellite) in the SAM, it no longer experiences any new cooling from the CGM, since the satellite hot haloes are assumed to be stripped instantaneously and added to the CGM of the central/host halo. Thus, satellites in some SAMs may tend to be ‘overquenched’, i.e. they may cease forming stars earlier than satellites of the same mass and host mass in hydrodynamic simulations. An additional secondary effect comes from the difference in the halo finders used to define haloes in both samples – TNG uses FoF, while SC-SAM uses ROCKSTAR (see Section 2.1). Haloes identified as two separate haloes with ROCKSTAR can often be considered to be a single FoF halo, which affects the halo mass function, the response to halo assembly bias, and the classification of dark-matter structure into haloes and subhaloes. These effects are considered in Appendix A, where we demonstrate that they do not influence the main conclusions in this work. We also refer the reader to Gómez et al. (2021), which examine the effect of using different merger tree building algorithms on the galaxy properties predicted by a SAM. While they find dependence on the halo finder employed and the resolution of the simulation when it comes to predicting the number of satellites, they conclude that the

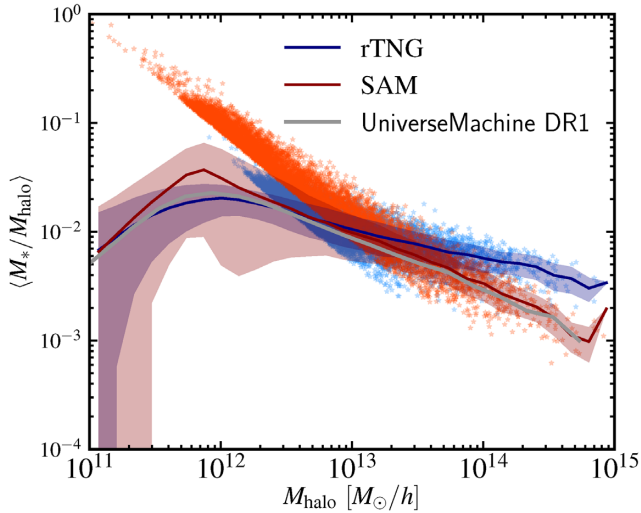


Figure 1. Stellar-to-halo mass ratio for all galaxies in TNG and SC-SAM (blue and red lines and shaded areas) as well as for the two stellar mass selected samples defined in the text (blue and red dots; see Section 2.3). The solid lines show the mean of the relation, while the shaded regions mark one standard deviation above and below the mean. In solid grey, we show the UniverseMachine (UM) DR1 prediction and find that while the rescaled TNG (rTNG) galaxies provide a very good match for smaller haloes, $M_{\text{halo}} \lesssim 10^{13} M_\odot h^{-1}$, on the high mass end, the TNG centrals appear to overproduce stars. On the other hand, SAM galaxies living in haloes with masses $M_{\text{halo}} \gtrsim 10^{13} M_\odot h^{-1}$ exhibit a better agreement with UM, but show a discrepancy around the peak, $M_{\text{halo}} \sim 10^{12} M_\odot h^{-1}$.

overall distributions of star formation rates, and stellar and gas masses agree well between different tree builders.

In Fig. 1, we show the stellar-to-halo mass ratio of central galaxies in all haloes for both models. The stellar masses output by TNG are rescaled so as to match the known outcome of the TNG300 model at the superior resolution of the TNG100 simulation (see Pillepich et al. 2018b, for more details). The rescaling of the stellar masses is performed only for this comparison plot and not throughout the paper. Relative to TNG, the SC-SAM produces lower values of m_*/M_{halo} at halo masses above ($M_{\text{halo}} \lesssim 10^{13} M_\odot h^{-1}$). In addition, the SC-SAM produces a sharper peak in the m_*/M_{halo} relation, and has considerably more scatter in this relation around its peak (see also Gabrielpillai et al., in preparation). As a result, the stellar mass (number density) selected SC SAM sample has a significantly larger number of galaxies occupying smaller haloes than the TNG sample. We conjecture this is due to a difference in the stellar feedback prescription in the two models, where TNG ties the effect to the dispersion velocity of haloes, while SC-SAM ties it to the maximum circular velocity (see Section 4). An additional effect we account for is the difference in the halo mass functions of the two finders ROCKSTAR (used in SC-SAM) and FoF (used in TNG), which affects the denominator of the m_*/M_{halo} relation (see e.g. Gómez et al. 2021, for a detailed discussion). As seen in Appendix A, ROCKSTAR finds 15 percent more haloes at the mass scale of $M_{\text{halo}} = 10^{12} - 10^{13} M_\odot h^{-1}$ compared with FoF. To better compare the halo masses of the two catalogues, we abundance match the haloes in ROCKSTAR to those in FoF. We show the average stellar-to-halo mass ratio for the entire population of subhaloes and compare them against the result from the UniverseMachine (UM) Data Release 1 (Behroozi et al. 2019), which is a semi-empirical model that is designed to match the observed stellar mass function. We find that SC-SAM is in very good agreement with UM for large halo masses,

whereas the rescaled TNG (rTNG; see Pillepich et al. 2018b, for details) run provides a better match for intermediate halo masses around the peak in the stellar-to-halo mass relation (see e.g. Ayromlou et al. 2021a; Gabrielpillai et al., in preparation, for an object-by-object comparison of the stellar-to-halo mass ratio in TNG and a SAM). We note that using a different stellar mass definition for the TNG galaxies (e.g. within 30 pkpc as in Pillepich et al. 2018b) would result in different stellar-to-halo mass ratio.

2.4 Assembly bias parameters

In this section, we review the definitions of the various halo parameters employed in this study (for more details, see Hadzhiyska et al. 2020). Appendix B discusses the correlations between the various parameters in addition to the halo assembly bias measured for concentration and environment.

2.4.1 Halo environment

We adopt the following definition of adaptive halo environment to assess its effects on the large-scale galaxy distribution:

1. Evaluate the dark matter overdensity field, $\delta(\mathbf{x})$, using triangular-shaped-cloud (TSC) interpolation on a 512^3 cubic lattice. In TSC, the fraction of a particle's mass assigned to a given cell of size Δx along one dimension is determined by the shape function: $S(x) = (1 - |x|/\Delta x)/\Delta x$. Each cell has size of $\Delta x = 205/512 \text{ Mpc } h^{-1} \approx 0.4 \text{ Mpc } h^{-1}$.
2. Smooth the density field with a Gaussian kernel on scales $R_{\text{smooth}} = 1.1, 1.5, 2.0, 2.5 \text{ Mpc } h^{-1}$.
3. The smoothing scale used for a given halo is selected based on the halo virial radius: i.e. we choose the closest smoothing scale to the halo radius rounding up. Thus, we make sure to always define the environment of a given halo on scales larger than its virial radius.
4. The halo local environment is determined by the value of the smoothed overdensity field interpolated at the halo centre-of-potential.

In previous works, we found that conditioning on this parameter in TNG leads to a substantial increase in the galaxy clustering on large scales (Hadzhiyska et al. 2020, 2021a). We conjectured that at a fixed mass, haloes residing in denser regions contain more galaxies on average than haloes in underdense regions due to experiencing more mergers and having a larger gas reservoir with which to form stars (Abbas & Sheth 2007; Pujol et al. 2017; Paranjape, Hahn & Sheth 2018; Shi & Sheth 2018). In Appendix B, we examine the correlation between environment and various other halo parameters. We find: on the high halo-mass end, the strongest correlation is between environment and the ratio M_{FoF}/M_{200} ; on the low-mass end, it is between environment and velocity anisotropy, defined as $1 - 0.5\sigma_{\text{tan}}/\sigma_{\text{rad}}$, which were previously found to be good predictors of galaxy clustering. In Fig. B2, we also present the halo clustering dependence on environment for both ROCKSTAR and FoF, confirming that haloes living in dense environments are substantially more clustered compared with haloes in underdense regions. This result is more noticeable for the ROCKSTAR haloes that tend to be better-deblended than the FoF ones, especially at these higher densities.

2.4.2 Halo concentration

The link between halo concentration and accretion history has been studied extensively in the literature (Navarro, Frenk & White 1997;

Wechsler et al. 2002; Ludlow et al. 2014, 2016). Very massive haloes with low concentration tend to have their subhaloes more spatially spread out as a result of having undergone a larger number of recent mergers (Bose et al. 2019). Therefore these haloes may contribute to a boost in the number of objects, and hence clustering, at scales comparable to the virial radius of these haloes. The effect of halo concentration on halo clustering and the relationship between halo concentration and halo mass have been thoroughly explored (Bullock et al. 2001a; Dutton & Macciò 2014; Ludlow et al. 2014; Diemer & Kravtsov 2015; Mao, Zentner & Wechsler 2018).

To obtain the halo concentration, we use the ROCKSTAR default definition, c_{NFW} , which follows the Navarro–Frenk–White prescription (Navarro et al. 1996), assuming that the density profiles of dark matter haloes are well-fitted by the formula

$$\rho_{\text{NFW}}(r) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2}, \quad (5)$$

where ρ_s is the characteristic density and r_s is the scale radius defined as

$$r_s \equiv \frac{r_{\text{vir}}}{c}, \quad (6)$$

with r_{vir} being the virial radius of the halo. We define this as the distance from the halo centre to the outer boundary within which the mean density is Δ_c times the critical density (see Section 2.1 for details on Δ_c).

Fig. B2, showing the halo assembly bias due to concentration, confirms the claim that less concentrated high-mass haloes exhibit an excess in clustering, whereas less concentrated low-mass haloes tend to be less clustered than their highly concentrated counterparts. Interestingly, we find very weak correlation between halo concentration and environment as defined in this study (see Fig. B1). We confirm the expected relationship between formation epoch and concentration.

2.4.3 Halo spin

Another halo parameter that we explore is halo spin, λ . This provides a measure of the angular momentum acquired by the halo. Its dependence on other halo properties, such as concentration and formation epoch, has been well studied (Bullock et al. 2001b; Bett et al. 2007; Rodríguez-Puebla et al. 2016; Johnson et al. 2019). The measurement of this parameter is quite sensitive to the particle resolution (the smaller the number of particles in a halo, the larger the error; Oñorbe et al. 2014; Benson 2017). In this work, we use the direct output from ROCKSTAR, which defines halo spin as Peebles (1969):

$$\lambda = \frac{J\sqrt{|E|}}{GM_{\text{vir}}^{2.5}}, \quad (7)$$

where J is the magnitude of the angular momentum of the halo and E is the total energy of the halo. We show the correlation between spin and environment, and spin and concentration in Fig. B1.

2.4.4 Properties computed from merger trees

By computing select quantities directly from the ROCKSTAR merger trees and using the available bijective matches between ROCKSTAR and FoF, we can examine the effects in both. These quantities are listed as follows:

(i) z_{form} : the time at which the halo first acquires 50 per cent of its final virial mass. The characteristic formation epoch of a halo is a direct indicator of its past accretion history at fixed present-day mass.

This quantity has a well-known connection to the halo concentration (which we recover in Fig. B1). Namely, the central regions of early-forming haloes collapse when the mean density of the Universe is higher and these haloes are thus more concentrated.

(ii) M_{peak} : the total mass of the particles within the virial radius of the halo (see Section 2.1 for definition of virial mass) at the time when it achieves its peak mass. A growing body of evidence seems to suggest that populating haloes based on their early-history properties (e.g. mass or velocity at time of infall or at their peak value) leads to a better agreement with observed galaxy clustering (e.g. Chaves-Montero et al. 2016). Haloes that have flown by a large object are well-known to be stripped of the outer layers of their dark matter component, retaining the baryonic core component. In such situations, the peak/infall mass acts as a better marker of their galaxy properties.

(iii) V_{peak} : the maximum value of V_{max} attained by the halo throughout its history, where V_{max} is the maximum value of the spherically averaged rotation curve for a given halo as a function of radius at some time epoch. Velocity-related parameters are thought to be more resilient to well-known effects, such as tidal stripping, satellite disruption and gas expulsion (in the form of AGN feedback, supernova feedback, stellar winds, etc.), when compared with mass-related ones (e.g. Guo et al. 2016a). This suggests they are less prone to change during the galaxy formation process and therefore might offer a better proxy for galaxy occupation.

3 RESULTS

In this section, we present the main results of the study. In particular, we present comparisons between our mass-selected SC-SAM and TNG samples for various galaxy–halo statistical quantities, such as occupancy distributions, two-point clustering, the galaxy assembly bias signal, and additional probes describing the galaxy distribution. Furthermore, we explore the dependence of halo occupancy and gas content on the halo environment, studying the differences between the two models.

3.1 Halo occupation distribution

The occupancy of haloes at fixed mass, also known as HOD, is a standard statistical quantity used to examine the galaxy–halo connection. It provides information about the average number of galaxies as a function of halo mass and can be broken down into the respective contributions from central and satellite galaxies. Its shape, in the case of an LRG sample, has been studied extensively and is well approximated by the empirical formula given in Zheng et al. (2005). One of the most common applications of HODs is in the construction of mock catalogues, i.e. galaxy catalogues meant to mimic the distribution of objects found in cosmological surveys. These mocks often assume a very simple relationship between haloes and the galaxies within them, motivated by the expected HOD of the population, in order to satisfy the hefty demands for volume and number of mock realizations (Baugh 2008).

In Fig. 2, we show the halo occupation distribution as measured in the SAM and TNG, split into centrals and satellites for the three mass-selected samples considered, with number densities of $n_{\text{gal}} = 0.0005, 0.001, \text{ and } 0.002 \text{ [Mpc } h^{-1}]^{-3}$. First focusing on the centrals, we see the SC-SAM sample has non-zero occupancy down to lower halo masses than TNG. Strikingly, the transition from 0 to 1 is markedly less steep in the case of the SC-SAM sample, where haloes are guaranteed a central (i.e. $\langle N_{\text{cen}} \rangle \rightarrow 1$) only after they reach a minimum mass of $M_{\text{halo}} \gtrsim 2 \times 10^{13} M_{\odot} h^{-1}$. On the other hand,

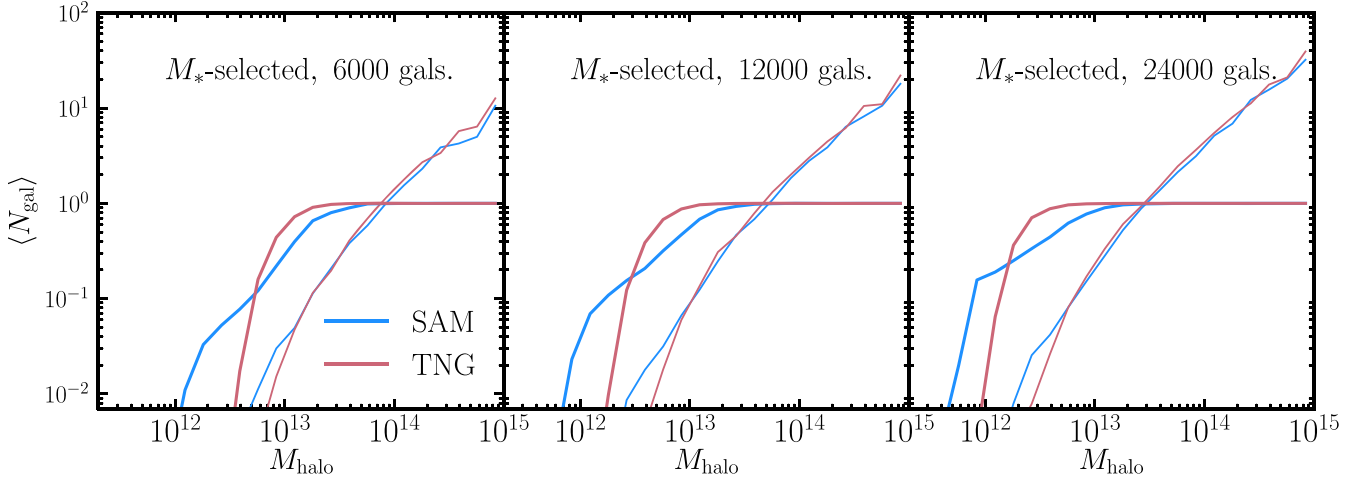


Figure 2. Halo occupation distribution for a stellar-mass-selected sample with number densities of $n_{\text{gal}} = 0.0005, 0.001$, and $0.002 \text{ [Mpc } h^{-1}]^{-3}$ (corresponding to 6000, 12 000, and 24 000 galaxies, respectively) at $z = 0$. In blue, we show the HOD of the SAM galaxy sample, while in red, we show the result for TNG. The galaxies are split into centrals (thick solid lines) and satellites (thin solid lines). The agreement between the TNG and SC-SAM satellite average occupation is very good for all panels. For all three number densities, the average HOD of the TNG centrals displays a sharper transition from 0 to 1 relative to the SC-SAM centrals. This is due to the tendency of SC-SAM to ‘paint’ galaxies on to lower mass haloes.

the average number of satellites per halo in the SAM displays very good agreement with TNG for all three samples. Although the HOD shape depends on the particular mass definitions adopted, changes to definitions only contribute a minor horizontal shift to the curves and do not alter the conclusions of our findings. Another factor is the different halo finding algorithms employed in both samples, but this effect is small and also does not affect our conclusions (see Fig. A2).

Fig. 3 demonstrates the dependence of the HOD on various halo parameters at $z = 0$ for a mass-selected sample with $n_{\text{gal}} = 0.001 \text{ [Mpc } h^{-1}]^{-3}$, corresponding to a lower mass threshold of $M_{\text{gal}} \approx 5 \times 10^{10} M_{\odot} h^{-1}$. For each halo mass bin, we order the haloes based on a given property and then select the top and bottom 25 per cent of haloes based on this ranking. We compute the average occupation in each mass bin and show the results for the following six halo parameters: environment, concentration (c_{NFW}), spin, peak circular velocity (V_{peak}), peak mass (M_{peak}), and formation epoch (z_{form}) (for more details, see Section 2.4). Qualitatively, the response for all of these parameters seems to be similar between the two models, confirming the findings of previous analyses, such as the typical inversion of the satellite and central occupation in the case of concentration and formation time. There are two noteworthy differences. The first is found in the case of the halo environment parameter, towards which the SC-SAM model appears to be less sensitive than TNG (especially for the satellites). The second is the stronger response to the following parameters in the case of SC-SAM: concentration, formation epoch, and to a lesser extent peak velocity. In the next section, we study how conditioning on each of these parameters manifests itself on the galaxy clustering.

3.2 Galaxy clustering

The spatial two-point correlation function, $\xi(r)$, measures the excess probability of finding a pair of objects at a given separation, r , with respect to a random distribution. It is a vital tool in cosmology for exploring the 3D distribution of any cosmological tracer and for constraining cosmological parameters. We compute the galaxy two-point correlation function using the natural estimator (Peebles 1980), where $DD(r)$ and $RR(r)$ are the number of data–data and random–

random pairs as a function of distance, respectively, as:

$$\hat{\xi}(r) = \frac{DD(r)}{RR(r)} - 1 \quad (8)$$

via the package CORRFUNC (Sinha & Garrison 2020), assuming periodic boundary conditions in a box of size $L_{\text{box}} = 205 \text{ Mpc } h^{-1}$. We estimate the uncertainties of the correlation function using jackknife resampling (Norberg et al. 2009). This is computed through dividing the simulation volume into 27 equally sized boxes before calculating the mean and jackknife errors on the correlation functions via the standard equations:

$$\bar{\xi}(r) = \frac{1}{n} \sum_{i=1}^n \xi_i(r) \quad (9)$$

$$\text{Var}[\xi(r)] = \frac{n-1}{n} \sum_{i=1}^n (\xi_i(r) - \bar{\xi}(r))^2, \quad (10)$$

where $n = 27$ and $\xi_i(r)$ is the correlation function value at distance r for subsample i (i.e. excluding the galaxies residing within volume element i in the correlation function computation). Additionally, we use jackknifing to estimate the error bars on ratios of correlation functions by first calculating the desired quantity in all $n_{\text{gal}} = 0.001 \text{ [Mpc } h^{-1}]^{-3}$ subsamples, then computing the mean and standard deviation as shown above. In this way, we diminish some of the effects of sample variance and provide a way of quantifying the significance with which we detect deviations from unity.

In Fig. 4, we show the two-point correlation function, $\xi(r) r^2$, for the SAM and SC-SAM mass-selected samples at galaxy number density of $n_{\text{gal}} = 0.001 \text{ [Mpc } h^{-1}]^{-3}$. In the upper panel, the two curves exhibit an excellent match on large scales (i.e. beyond the one-halo term, $\sim 2 \text{ Mpc } h^{-1}$); on intermediate scales ($\sim 1 \text{ Mpc } h^{-1}$) in the lower panel (and less so in the upper panel), the SC-SAM sample appears to be more strongly clustered. The behaviour can be attributed to the following factors. First, SC-SAM tends to populate haloes of lower mass, which have smaller radii, so the exclusion radius, marking the transition between the one- and two-halo terms, manifests itself at smaller scales. The slightly stronger correlation signal at $r \sim 10 \text{ Mpc } h^{-1}$ of the SC-SAM centrals can be explained by the fact that SC-SAM preferentially populates highly concentrated

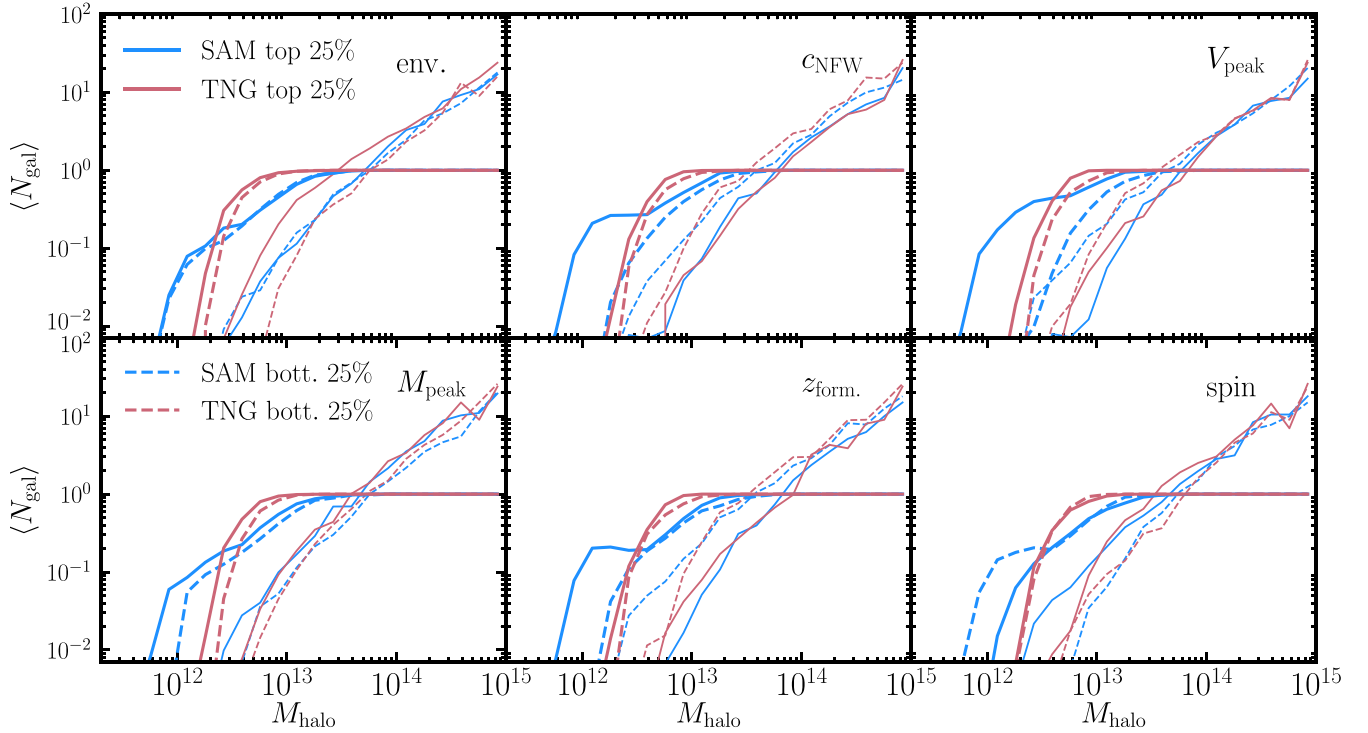


Figure 3. Dependence on the halo parameters of the halo occupation distribution for a stellar-mass-selected sample with a galaxy number density of $n_{\text{gal}} = 0.001 \text{ [Mpc } h^{-1}]^{-3}$ (corresponding to 12 000 galaxies) at $z = 0$. In blue, we show the HOD of the SAM galaxy sample, while in red, we show the result for TNG. The galaxies are split into centrals (thick lines) and satellites (thin lines). In each mass bin, we select the top and bottom 25 per cent of haloes (solid and dashed lines, respectively) ordered by one of six properties: environment, concentration, formation epoch, spin, peak circular velocity, and peak mass (see Section 2.4). The qualitative response to the various parameters is similar in both models, with slight discrepancies. For example, the environmental effect is much stronger for TNG satellites and so is the response of the SC-SAM to V_{peak} , z_{form} , and c_{NFW} . The latter is likely due to the explicit appearance of these parameters in the model and/or the strong correlation between some of them (see Section 2.1 and Appendix B).

haloes at fixed mass which at this mass-scale, exhibit strong halo assembly bias (see Fig. B2). On small scales ($0.1 \text{ Mpc } h^{-1} \lesssim r \lesssim 1 \text{ Mpc } h^{-1}$), the satellite distribution of SC-SAM appears to be shifted to slightly larger scales. We conjecture that this is because the SAM satellites occupy slightly less concentrated haloes as seen in Fig. 2 (see Section 2.1 for details on the SC-SAM model) with a more spread-out satellite distribution, which shifts outward the one-halo clustering signal ($r \sim 1 \text{ Mpc } h^{-1}$). Another noticeable feature is the larger size of the error bars on the upper panel as a result of the larger shot noise of the satellites compared with the centrals. The error bars indicate the amount of variance resulting from the limited simulation volume, but the relative agreement or disagreement of the two models cannot be interpreted based on those, as the SAM and TNG results are calculated based on the same underlying halo distribution.

3.3 Galaxy assembly bias signature

Galaxy assembly bias can be thought of as a ‘convolution’ between two separate effects. It is a result of halo assembly bias and occupancy variation, where the former is the dependence of the clustering of DM haloes on parameters other than mass, while the latter refers to the dependence of halo occupancy on halo parameters other than mass. A standard way to measure galaxy assembly bias is by comparing the two-point-correlation function from a given sample with a sample in which the galaxies populating haloes are randomly reassigned to haloes within the same mass bin (Croton et al. 2007). This ‘shuffling’ eliminates the dependence of halo occupancy on any secondary properties other than halo mass. If the ratio of the

two-point correlation function of the shuffled to unshuffled sample differs from zero, then the galaxy assembly bias signal is non-zero. This method of defining galaxy assembly bias is only meaningfully measurable on large scales. Therefore, we opt to preserve the relative positions of all galaxies in the FoF halo with respect to their host halo centre when transferring them to their newly assigned host. In this way, the shuffled sample maintains the same one-halo term as the original sample, and the comparison is purely between the two-halo term contribution to both.

In Fig. 5, we illustrate the galaxy assembly bias signal for the three mass-selected samples considered in this study, with number densities of $n_{\text{gal}} = 0.0005, 0.001$, and $0.002 \text{ [Mpc } h^{-1}]^{-3}$. We shuffle the halo occupation number in mass bins of width $\Delta \log(M_{\text{halo}}) = 0.2$ dex. For $r \gtrsim 1 \text{ Mpc } h^{-1}$, we find that the ratio deviates from one on average by about 15 per cent, 18 per cent, and 20 per cent for the three samples, respectively. Moreover, it is noteworthy that the SC-SAM and TNG samples appear in very good agreement with each other, with the largest difference between the two being found in the sample with highest number density. On small scales ($r \lesssim 1 \text{ Mpc } h^{-1}$), the ratio is consistent with one, as we preserve the one-halo term when performing the shuffling. This finding suggests the SC-SAM model naturally induces a non-negligible galaxy assembly bias signal, which is similar to that seen in a hydrodynamical simulation.

Studying the response of both models to various halo properties is important for understanding the source of the assembly bias signal. Instead of randomly shuffling the halo occupancies in bins of halo mass, we now sort the haloes by a selected property: environment, concentration, peak velocity, peak mass, formation epoch and spin, in

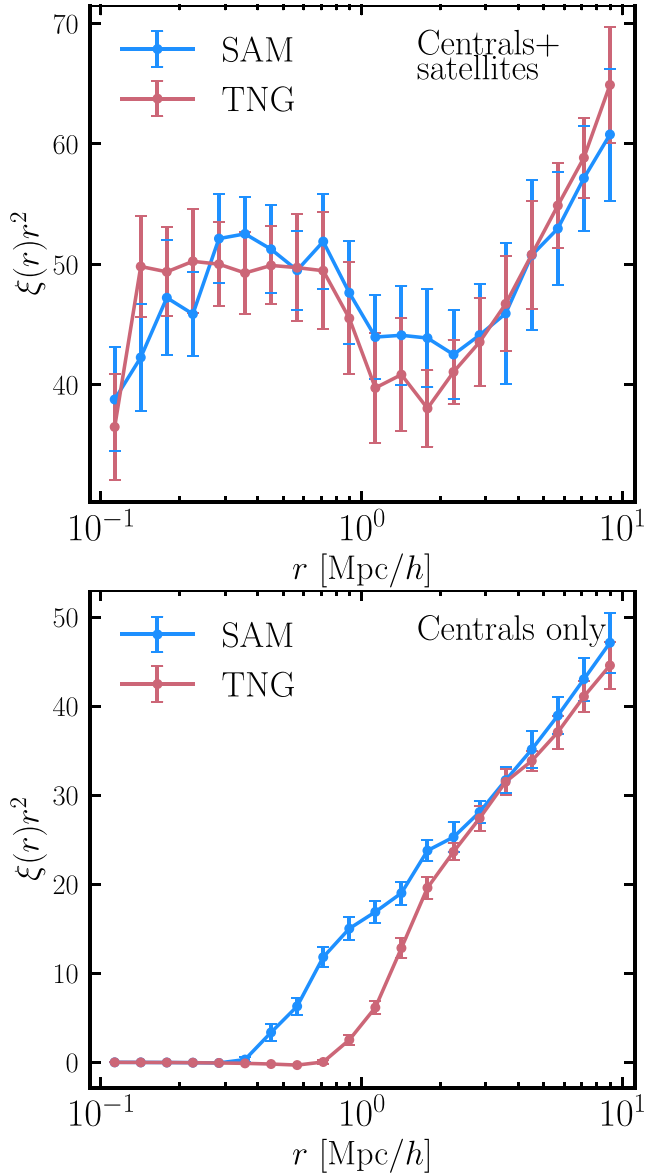


Figure 4. Correlation function for a stellar-mass-selected sample with a galaxy number density of $n_{\text{gal}} = 0.001 \text{ [Mpc } h^{-1}]^{-3}$ (corresponding to 12 000 galaxies) at $z = 0$. In blue, we show the two-point correlation function of the SAM galaxy sample, while in red, we show the result for TNG. The bottom panel shows the two-point clustering of only the centrals (roughly 9000 objects) for both samples. The error bars are computed using jackknifing. We see that on large scales, the agreement is excellent. On intermediate scales, centrals are slightly more clustered in SC-SAM, due to the tendency of SC-SAM to populate haloes of lower mass with smaller radii, thus reducing the range of the one-halo term. On small scales, SC-SAM exhibits higher clustering.

either descending or ascending order, and then distribute the galaxy occupations to the sorted haloes in descending order. The result of this exercise is shown in Fig. 6. The strongest response of the clustering is to halo environment. We conjecture that it is predominantly the result of halo assembly bias, which explains the stronger signal for the SC-SAM sample. On one hand, as a result of the flatter HOD shape of the SC-SAM centrals (see Fig. 2), the reordering procedure employed when measuring galaxy assembly bias preferably picks haloes in dense environments across a wider mass scale compared with TNG. This inevitably boosts the clustering signal as these haloes

are significantly more biased. To test this hypothesis, we create an ad-hoc SAM galaxy sample assuming the HOD shape of TNG and find that the assembly bias signature becomes nearly identical to that of TNG with differences comparable to the halo assembly bias signal seen in Fig. B2. On the other hand, a secondary effect comes from the fact that SC-SAM uses the ROCKSTAR halo catalogue which exhibits a slightly stronger dependence on environment than FoF (see Fig. B2). Occupancy variation also plays a role in determining galaxy assembly bias: the response to environment is similar in both, with TNG exhibiting a slightly stronger dependence (see Fig. 3). Another interesting observation, which echoes our finding from Fig. 3, is that reordering the haloes in terms of their concentration, circular velocity, formation epoch, and peak mass results in a significant increase in the SC-SAM clustering, whereas TNG is less sensitive to these parameters (a similar conclusion was reached in Hadzhiyska et al. 2020). This makes sense in light of the similarly stronger dependence of the halo occupancy on these parameters in the SAM.

3.4 Galaxies and their gaseous environment

One of the major differences between SAMs and numerical hydrodynamic simulations is that the numerical simulations explicitly simulate the distribution of gas on both large and small scales, and can therefore make detailed predictions about the temperature and density of this gas, as well as how these properties are affected by the baryonic feedback processes such as stellar and AGN feedback. SAMs make simplifying assumptions, such as that halo gas is in an isothermal sphere with a fixed density profile, and cannot make any predictions for how gas is distributed outside of haloes. The availability of gas is expected to have an important affect on observable galaxy properties. In order to interpret our results, it is therefore interesting to explore whether the way in which gas traces the underlying dark matter shows any dependence on environment, and whether this is the same in the two models.

One might intuitively expect to find that objects lying in high-density regions have a higher gas content and vice versa for haloes in low-density environments. To test this conjecture, we show the ratio between gas and dark-matter content in TNG and SAM haloes as a function of redshift in Fig. 7. We select abundance-matched haloes with $\log M_{\text{halo}} = 13.0\text{--}13.1$ from the two catalogues. We obtain the abundance matched samples by rank-ordering the haloes by halo mass in both catalogues. Then we find the halo indices corresponding to the mass range of interest in the FoF catalogue and select the haloes from the ROCKSTAR catalogue based on those.

This scale roughly corresponds to the average host halo mass for the stellar-mass-selected galaxy sample presented in this study. We then choose the top and bottom 10 per cent of haloes based on the value of their environment parameter, resulting in four sets of 50 haloes each. In the case of SC-SAM, we do not resolve the individual gas particles/cells, so the ratio between the total amount of hot gas in the parent halo and its virial dark-matter mass is shown instead. In the case of TNG, we further show the differential gas mass (sum of the masses of all gas cells) within shells surrounding the central galaxy. We adopt the following inner and outer radii pairs: $(r_{\text{inner}}, r_{\text{outer}}) = \{(0, 1/2 \times r_{\text{vir}}), (r_{\text{half}}, 1.2 \times r_{\text{vir}})\}$, where r_{half} is the stellar half-mass radius for the central subhalo reported in TNG and r_{vir} is the virial radius (‘Group_R_TopHat200’). These are chosen to be at the scale of the typical galaxy formation processes, to get a sense of the gas reservoir available locally for forming stars and accreting on to the central galaxy. We see that the SAM result appears to be almost completely insensitive to environmental variations. On the other hand, the amount of gas surrounding the TNG centrals

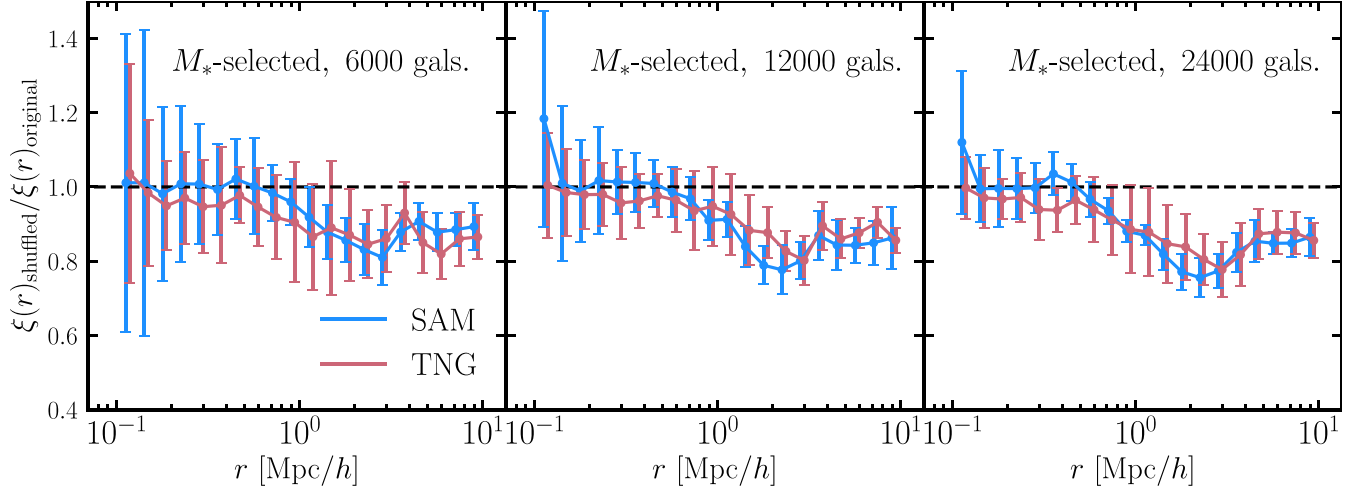


Figure 5. Galaxy assembly bias of a stellar-mass-selected sample with number densities of $n_{\text{gal}} = 0.0005, 0.001, \text{ and } 0.002 \text{ [Mpc } h^{-1}]^{-3}$ (corresponding to 6000, 12 000, and 24 000 galaxies, respectively) at $z = 0$. The signal for each model is computed by randomly shuffling the halo occupations at fixed halo mass, while preserving the one-halo term. In blue, we show the galaxy assembly bias signal of the SAM galaxy sample, while in red, we show the result for TNG. The error bars are computed using jackknifing of the ratios. A deviation from unity suggests that a mass-only HOD is incapable of matching the clustering of LRG-like galaxies. The signal is very similar for both models in the first two panels, but shows a slightly larger discrepancy in the third, where the SC-SAM galaxies display even stronger galaxy assembly bias.

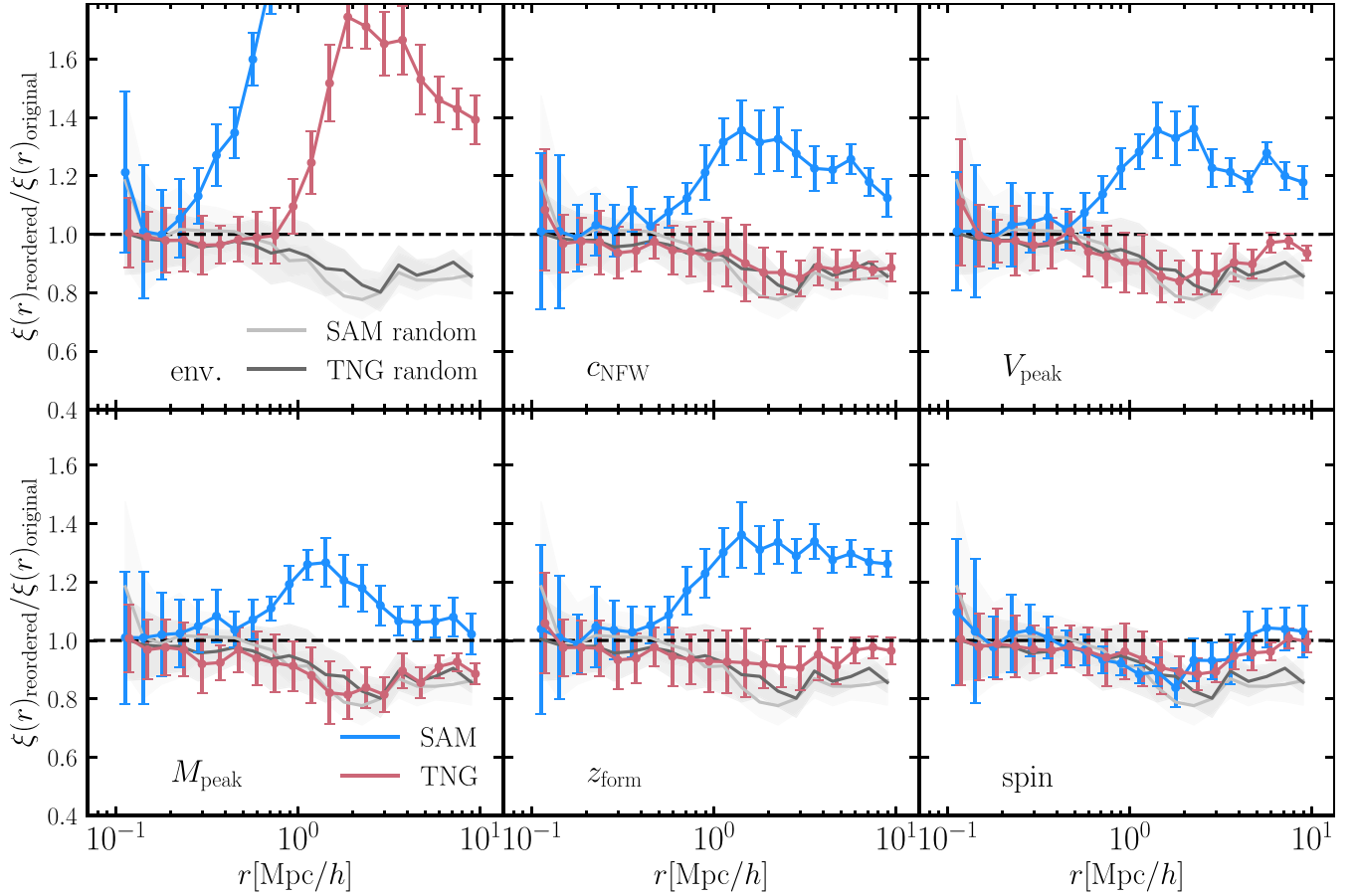


Figure 6. Dependence on different halo parameters of the galaxy assembly bias for a stellar-mass-selected sample with a galaxy number density of $n_{\text{gal}} = 0.001 \text{ [Mpc } h^{-1}]^{-3}$ (corresponding to 12 000 galaxies) at $z = 0$. In blue, we show the galaxy assembly bias signal of the SAM galaxy sample, while in red, we show the result for TNG. In grey and black, we show the SC-SAM and TNG curves from the middle panel of Fig. 5. Error bars are computed using jackknifing. The environmental dependence, which is significant in both models, is stronger in the SAM. This is likely because the halo assembly bias with respect to environment is stronger for the ROCKSTAR catalogue compared with the FoF catalogue. As seen in Fig. 3, the SC-SAM sample is very sensitive to the parameters: concentration, formation epoch, and peak circular velocity, all of which appear explicitly in the SAM model (see Section 2.1).

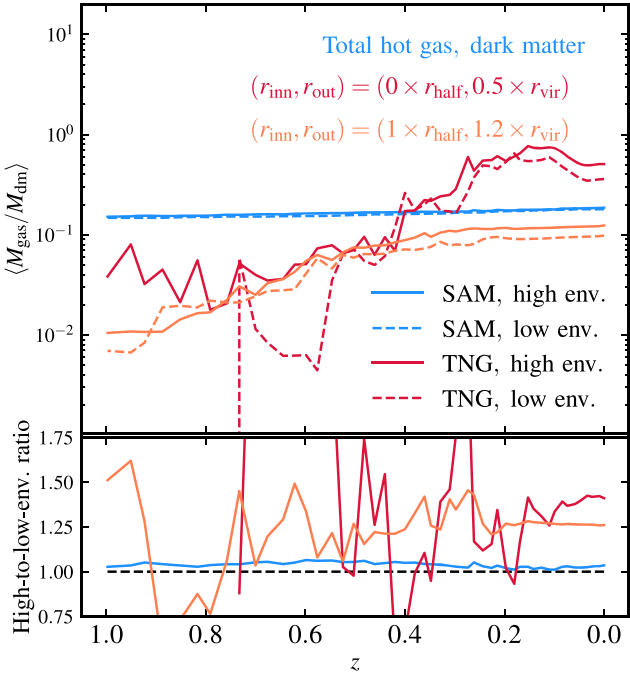


Figure 7. Ratio of the gas to dark matter mass for SC-SAM (blue) and TNG (red). In the case of TNG, these numbers come from summing the masses of all gas cells contained within an annulus with inner and outer radius pairs – $(0, 1/2 \times r_{\text{vir}})$ and $(r_{\text{half}}, 1.2 \times r_{\text{vir}})$ – as a function of redshift for $\log M = 13$ haloes belonging to high- (marked with solid lines) and low- (marked with dashed lines) density environments. In the case of SC-SAM, we take the ratio between the total amount of hot gas and the total dark matter mass, as the model does not output the individual gas profiles of haloes. We see that the gas content in TNG haloes exhibits a much stronger connection with environment; the effect is minimal in SC-SAM. The lower panel shows the ratio between the solid dashed curves for each of the three cases. Overall, the amount of gas in TNG demonstrates a stronger dependence on environment than the SC-SAM. In particular, near $z = 0$ (the time at which the haloes are selected), we see that the gas-to-dark-matter ratio is much more pronounced for TNG than SC-SAM, at 25 per cent and 40 per cent for the orange and red lines, respectively, compared with 2 per cent for the SAM. In TNG, gas is supplied to the halo through filaments and sheets and is thus inherently environment-dependent; in SC-SAM, the amount of gas injected into and ejected from the halo is determined by other halo properties (halo mass, concentration, maximum circular velocity).

exhibits a much more noticeable dependence on environment. We emphasize that our halo selection and environment are defined at $z = 0$. We can connect the findings of this section to some of the takeaways from Fig. 2: namely, we noted that the dependence of the average number of satellites per halo mass on environment was much weaker in SC-SAM compared with TNG. This is in agreement with the conjecture of Fig. 7 that haloes in higher density regions host a larger number of satellites due to their ex-situ introduction into the main halo through filaments (along with gas). There are additional environmental effects that go in the opposite direction of destroying satellites in dense environments such as ram pressure stripping and tidal interactions with the central, but these effects appear to be of secondary importance, as they do not have a visibly strong impact on the satellite occupancy (see Fig. 2) and quantifying their size is not the subject of this study. We make available time-lapse visualizations of the selected TNG haloes (the entire FoF group) here: <https://drive.google.com/drive/folders/1AZX8eocrKwHGxqtGSf5XioXhI5kFkswl?usp=sharing>.

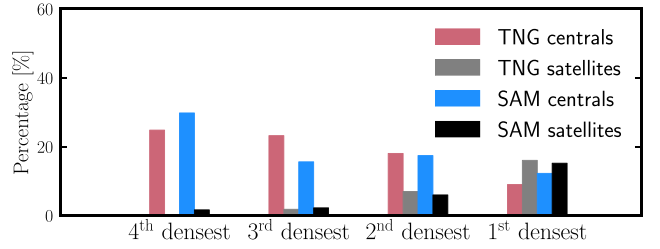


Figure 8. Percentage of galaxies found in the different density regions for SAM and TNG, split into centrals and satellites. The sum of the central and satellite fractions for each model (TNG and SC-SAM) in each of the four density regions contributes 25 per cent.

Another feature of Fig. 2 that might be affected by the absence of environmental dependence in the SAM is the central occupation. To test this, we examine the spatial distribution of the galaxies as a function of local environment. We order the galaxies in the TNG stellar-mass-selected sample by their local environment and split the sample into four subsamples of equal size (i.e. 3000 galaxies in each). Using the same threshold values of the local environment, we obtain four subsamples for SC-SAM and display the results in Fig. 8, separated into centrals and satellites. The most conspicuous discrepancy between the two models is in the densest and least dense regions (which have the most pronounced effect on the clustering because their galaxy bias displays the largest fluctuation). There is a substantially larger number of SC-SAM centrals (29.8 per cent) and satellites (1.6 per cent) inhabiting the least dense region in contrast to TNG (24.8 per cent and 0.2 per cent) and a somewhat larger number of SC-SAM centrals (12.2 per cent) residing in the densest regions compared with TNG (9.0 per cent). The third densest region compensates for the difference in central fractions with TNG containing 23.2 per cent of them while SC-SAM only 15.6 per cent. The larger number of SAM galaxies in the least dense region seems consistent with the tendency seen in Fig. 2 of the SAM to populate haloes of lower masses with galaxies. This effect, in conjunction with the larger fraction of SAM centrals in the densest region (with the highest bias), yields a consistent correlation function between the two samples (see Fig. 4).

3.5 Bias and correlation coefficient

In traditional cosmological studies of the matter distribution and its relationship to cosmological parameters, it is common to describe the relationship between observable tracers (galaxies) and the underlying mass distribution in terms of a ‘bias’ parameter. The galaxy autocorrelation function, $\xi_{gg}(r)$ is related to the matter correlation function, $\xi_{mm}(r)$, through the real-space galaxy bias, $\tilde{b}(r)$, in the following way:

$$\xi_{gg}(r) = \tilde{b}^2(r) \xi_{mm}(r). \quad (11)$$

Furthermore, one can study the galaxy-matter connection through their cross-correlation function, $\xi_{gm}(r)$, which can be related to ξ_{mm} through \tilde{b} and the cross-correlation coefficient, \tilde{r} (Hayashi & White 2008; Desjacques, Jeong & Schmidt 2018):

$$\xi_{gm}(r) = \tilde{b}(r) \tilde{r}(r) \xi_{mm}(r), \quad (12)$$

where the cross-correlation coefficient is given by:

$$\tilde{r}(r) = \frac{\xi_{gm}(r)}{[\xi_{gg}(r) \xi_{mm}(r)]^{1/2}}. \quad (13)$$

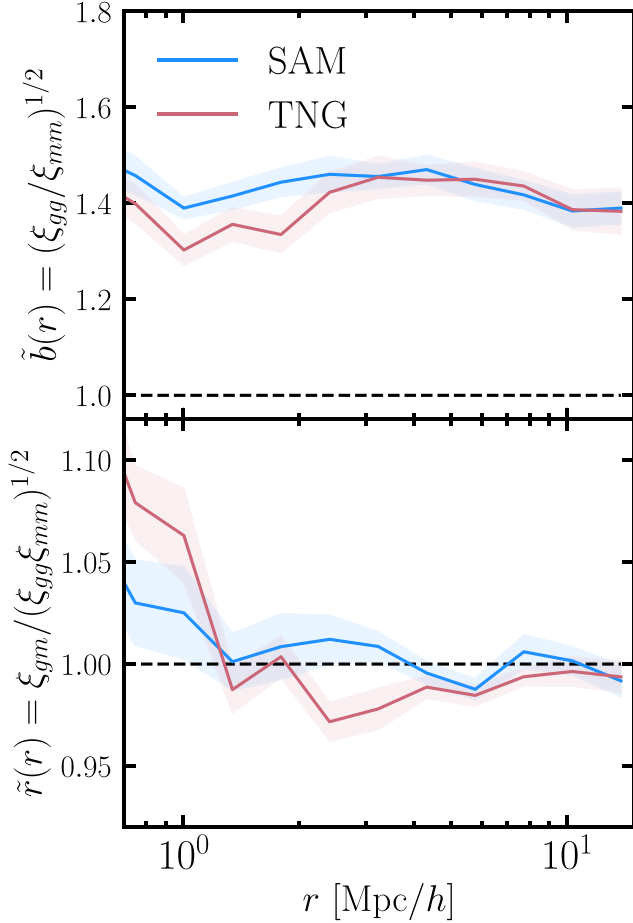


Figure 9. Bias and cross-correlation coefficient for a stellar-mass-selected sample with a galaxy number density of $n_{\text{gal}} = 0.001 [\text{Mpc } h^{-1}]^{-3}$ (corresponding to 12000 galaxies) at $z = 0$. In blue, we show the bias and correlation coefficient of the SAM galaxy sample, while in red, we show the result for TNG. Error bars are derived using jackknifing. The bias in both models show good agreement and approaches $\tilde{b}(r) \approx 1.4$ on large scales. The cross-correlation coefficient for TNG is lower on intermediate scales, suggesting the linear bias assumption is imperfect. In the case of SC-SAM, it is consistent with one.

The equations above are general and may be taken as definitions of the scale-dependent galaxy bias $\tilde{b}(r)$ and cross-correlation coefficient $\tilde{r}(r)$, although alternative definitions exist and are not equivalent. We note that the quantity $\tilde{r}(r)$ in real space is not constrained to be less than or equal to one, as is the case in Fourier space. However, one expects $\tilde{r}(r)$ to approach unity on large scales, where the observed correlation is sourced from the gravitational field of the total matter.

Fig. 9 shows the galaxy bias and the cross-correlation coefficient for a stellar-mass-selected sample with number density of $n_{\text{gal}} = 0.001 [\text{Mpc } h^{-1}]^{-3}$ for SC-SAM and TNG. The galaxy bias is consistent between the two models and levels out to about $\tilde{b}(r) \approx 1.4$ around $\sim 10 \text{ Mpc } h^{-1}$ at the onset of the linear bias regime. This result is expected, as the galaxy clustering matches well (Fig. 4) and the matter–matter correlation function is the same for both, being provided by dark matter particles in the simulation. The cross-correlation coefficient approaches one for $r \sim 1 \text{ Mpc } h^{-1}$, and on smaller scales, the two models show a moderate difference. In particular, for TNG, we find that $\tilde{r}(r) \approx 0.97$ for $r \approx 3 \text{ Mpc } h^{-1}$, suggesting that the linear bias approximation breaks down on these

scales. On the other hand, the cross-correlation coefficient of the SC-SAM sample is consistent with one.

3.6 Additional probes of the galaxy distribution

In this section, we explore alternative statistical probes of the large-scale galaxy distribution, such as weak lensing, galaxy-void cross-correlation, and cumulants of the smoothed galaxy density field. We pursue the goal of further describing the galaxy distribution of the TNG and SC-SAM samples in addition to documenting the differences between the two models.

3.6.1 Density field cumulants

The cumulants of the density field are measured from the moments of the smoothed galaxy overdensity field. They can be understood as degenerate N -point correlation functions or as integrated monopole moments of the bispectrum, which are closely related to neighbour counts both in their physical interpretation as well as in their algorithmic implementation. Of particular importance is the scale of galaxy clusters (~ 3 to $10 \text{ Mpc } h^{-1}$), where we expect that galaxy population methods could exhibit substantial differences (Bernardeau 1994; Gaztanaga 1994). The procedure we follow (similar to Hadzhiyska et al. 2021a) can be outlined as follows:

- (i) Divide the TNG box into 512^3 cubes of side $\sim 0.4 \text{ Mpc } h^{-1}$ and compute the counts-in-cell density field in each as $\delta_i = N_i/\bar{N} - 1$.
- (ii) Convolve it with a Gaussian filter of smoothing scale $R = \{3.75, 5, 6.25, 7.5, 8.75, 10\} \text{ Mpc } h^{-1}$ to get the smoothed density field δ_R .
- (iii) Compute the third moment of the density contrast by averaging the cubed overdensity over all cells, $\langle \delta_R^3 \rangle$.
- (iv) Study the change of the third moment values as a function of smoothing scale for different galaxy distributions.

In Fig. 4, we show that the two-point correlation functions of the SC-SAM and TNG galaxy populations exhibit very good agreement with each. However, this does not guarantee the consistency of higher order statistics. The averaged third moment of the galaxy density field, $\langle \delta_R^3 \rangle$, displayed in Fig. 10, offers a compressed view of the three-point correlation function. The upper panel shows $\langle \delta_R^3 \rangle$ as a function of smoothing scale. Increasing the smoothing scale R smooths out more and more fluctuations of the galaxy density field; thus, the value of $\langle \delta_R^3 \rangle$ decreases as expected. The lower panel shows the ratio of the SC-SAM curve to TNG. Surprisingly, we find that $\langle \delta_R^3 \rangle$ differs by ~ 20 percent despite the near perfect match of the two-point correlation functions beyond $r > 2 \text{ Mpc } h^{-1}$. This finding argues strongly in favour of including alternative statistical probes when we compare and calibrate galaxy population models. Recovering the two-point correlation function may not be a sufficient condition for generating realistic galaxy catalogues.

3.6.2 Galaxy–galaxy lensing

Galaxy–galaxy lensing, also known as stacked lensing, offers a unique window for mapping the underlying matter distribution by measuring the cross-correlation of large-scale structure tracers with the shapes of background galaxies. Stacked lensing measurements are expected to be one of the most powerful probes of modern cosmology, allowing cosmologists to address fundamental physics questions such as the nature of dark energy and dark matter and the mass hierarchy of neutrinos (e.g. Oguri & Takada 2011). Furthermore,

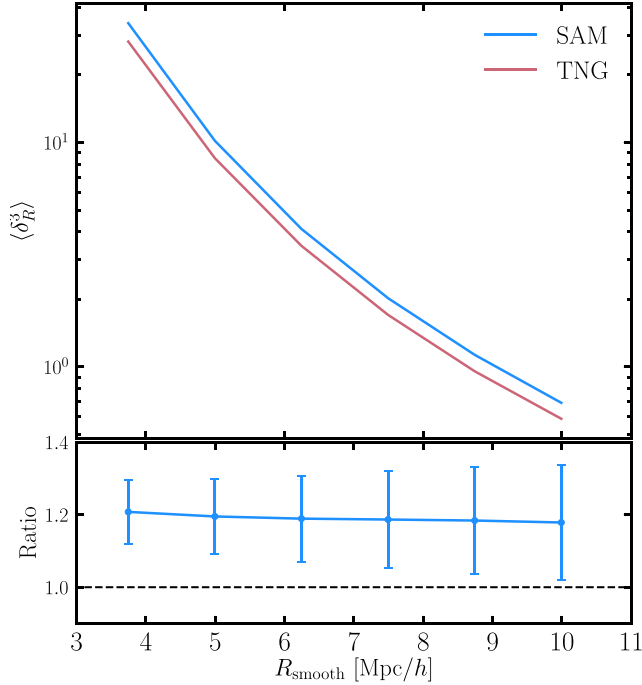


Figure 10. Cumulant of the smoothed galaxy density, $\langle \delta_R^3 \rangle$, as a function of smoothing scale for a stellar-mass-selected sample with a galaxy number density of $n_{\text{gal}} = 0.001 \text{ [Mpc } h^{-1}]^{-3}$ (corresponding to 12 000 galaxies) at $z = 0$. In blue, we show the cumulative density of the SAM galaxy sample, while in red, we show the result for TNG. $\langle \delta_R^3 \rangle$ decreases as we increase the smoothing. The lower panel, showing the ratio between the two curves, presents a ~ 20 per cent offset between the two models, despite the well-matched two-point clustering of the samples. This suggests that including higher point statistics in the analysis and calibration of galaxy formation models can help us differentiate between them and allows us to obtain more accurate depictions of the large-scale galaxy distribution.

by combining stacked lensing and autocorrelation measures of the same foreground galaxies, one can constrain cosmology by breaking degeneracies between galaxy bias and cosmological parameters (e.g. Seljak et al. 2005; Sunayama et al. 2020). Hydrodynamical simulations are invaluable in this manner: they allow us to compare the weak lensing signal, as predicted by various galaxy population prescriptions, and evaluate their ability to match the ‘true’ galaxy distribution, where ‘true’ is defined by the hydro simulation. Although the hydro model may not be fully consistent with observations, it is still a plausible model for how matter, baryonic and dark, is distributed in the Universe. It is, therefore, crucial to demonstrate that the galaxy population techniques we adopt when creating high-fidelity mock catalogues are flexible enough to recover various statistical properties of the galaxy sample in a hydro simulation.

As a measure of weak lensing, we consider the excess surface mass density profile, denoted as $\Delta \Sigma(r)$. It is obtained by first calculating

$$\Sigma(r_p) = \bar{\rho} \int_0^{\pi_{\text{max}}} \left[1 + \xi_{\text{gm}} \left(\sqrt{r_p^2 + \pi^2} \right) \right] d\pi, \quad (14)$$

where $\bar{\rho}$ is the mean matter density while r_p and π are the distances perpendicular and parallel to the line of sight, respectively. Then one can find the excess surface mass density as

$$\Delta \Sigma(r_p) = \bar{\Sigma}(< r_p) - \Sigma(r_p), \quad (15)$$

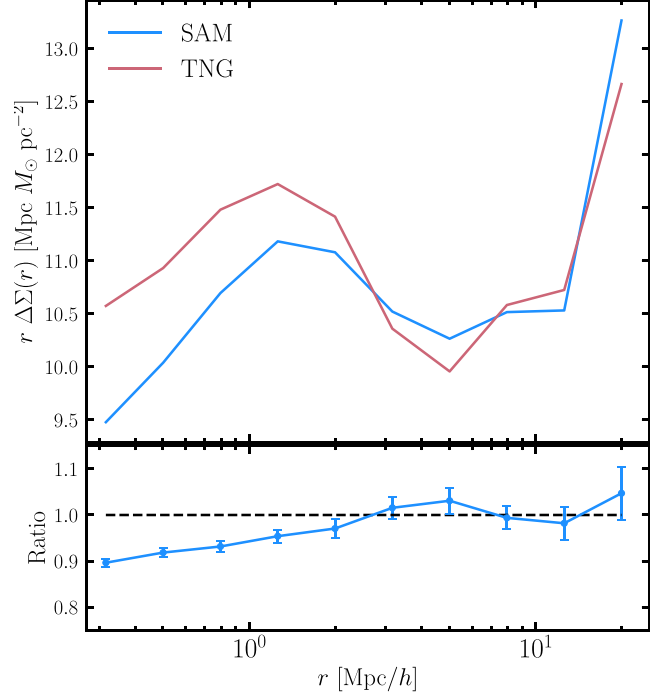


Figure 11. Galaxy–galaxy lensing for a stellar-mass-selected sample with a galaxy number density of $n_{\text{gal}} = 0.001 \text{ [Mpc } h^{-1}]^{-3}$ (corresponding to 12 000 galaxies) at $z = 0$. In blue, we show the weak lensing signal of the SAM galaxy sample, while in red, we show the result for TNG. On large scales, the weak lensing signal appears to be consistent between the two models. On small scales, there is a discrepancy consistent with Fig. 4.

where the mean surface mass density interior to the projected radius is given by

$$\bar{\Sigma}(< r_p) = \frac{1}{\pi r_p^2} \int_0^{r_p} \Sigma(r'_p) 2\pi r'_p dr'_p. \quad (16)$$

In Fig. 11, we show the galaxy–galaxy lensing signal of the SC-SAM and TNG stellar-mass-selected samples at $z = 0$. On scales larger than $r \gtrsim 3 \text{ Mpc } h^{-1}$, the two are in good agreement with each other, whereas on smaller scales, they deviate by ~ 10 per cent. This small-scale discrepancy should not come as a surprise given that their one-halo term contribution to the two-point clustering differs (see Fig. 4). The large scale consistency is encouraging, as it suggests that they have a similar connection, statistically speaking, to the underlying matter field.

3.6.3 Galaxy–void correlation function

The study of cosmic voids offers an exquisite opportunity to study cosmology and understand what our Universe is made of. Voids are huge low-density regions ($10\text{--}100 \text{ Mpc } h^{-1}$) that have undergone little non-linear growth (Gregory & Thompson 1978). Voids are complementary to both galaxy clustering and early-Universe measurements. They can help break existing degeneracies between cosmological parameters, as they are sensitive to a number of effects, such as redshift space distortions, baryon acoustic oscillations, neutrino effects, and the integrated Sachs–Wolfe effect (e.g. Kreisch et al. 2019). Cross-correlating the galaxy positions and the location of the voids offers insight into the relationship between large underdensities and densely clustered regions, exaggerating the most galaxy-deprived regions. It has the potential to differentiate between models that

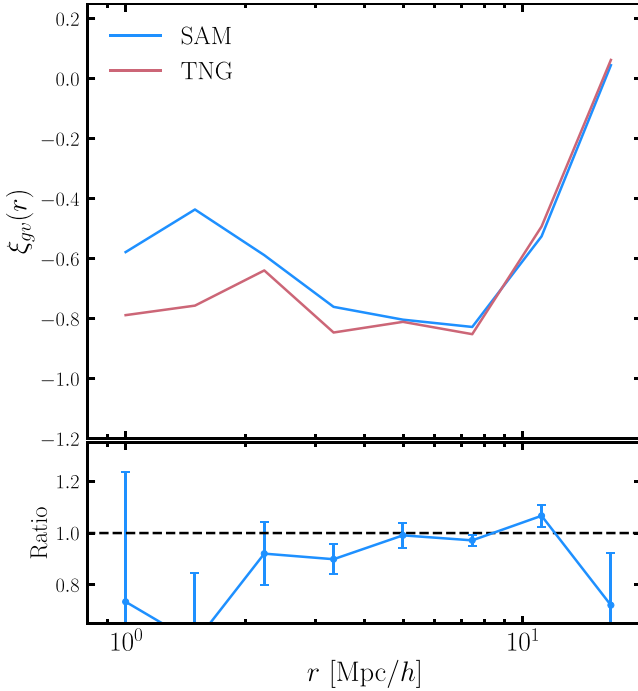


Figure 12. Galaxy–void correlation function for a stellar-mass-selected sample with a galaxy number density of $n_{\text{gal}} = 0.001 [\text{Mpc } h^{-1}]^{-3}$ (corresponding to 12 000 galaxies) at $z = 0$. In blue, we show the galaxy–void correlation function of the SAM galaxy sample, while in red, we show the result for TNG. On large scales, the two models show very good agreement (except for the last bin, which suffers from poorer statistics due to the limited size of the box), while on small scales, there is an evident discrepancy. Its significance is, however, difficult to assess due to the box limitations of the TNG simulation, so we do not place much importance on it.

exhibit the same two-point autocorrelation statistics and thus help us generate more accurate mock catalogues to match the full statistical properties of the galaxy population.

We use a simplified procedure for locating voids in the TNG simulation which was first outlined in Hadzhiyska et al. (2021a):

- (i) Divide the TNG box into 128^3 cubes of side $\sim 1.6 \text{ Mpc } h^{-1}$.
- (ii) Find the distance to the third nearest galaxy (i.e. the void sizes) measured from the centre of each cube (i.e. the void centres). As discussed in Hadzhiyska et al. (2021a), the particular choice of ‘third nearest’ does not affect qualitatively the void catalogues.
- (iii) Order the obtained void candidates based on their size in descending order and for each object in the list, remove all voids whose centres lie within the boundaries of that object.
- (iv) Record the void sizes and void centres for both the SAM and the TNG models.

We identify on the order of 2000 voids using the two galaxy samples for TNG and SC-SAM (see Section 2.3), the smallest of which have void sizes of $\sim 9 \text{ Mpc } h^{-1}$, while the largest reach $25 \text{ Mpc } h^{-1}$.

In Fig. 12, we show the galaxy–void cross-correlation function for the TNG and SC-SAM stellar-mass-selected samples with $n_{\text{gal}} = 0.001 [\text{Mpc } h^{-1}]^{-3}$ at $z = 0$. On these scales, the cross-correlation between galaxies and voids is weak and negative, in agreement with previous works (Hadzhiyska et al. 2021a). On scales of $r \gtrsim 3 \text{ Mpc } h^{-1}$, the ratio of the two is consistent with one, but deviates on smaller scales. Given that the average size of the voids we have identified is $\sim 10 \text{ Mpc } h^{-1}$, we do not place much importance on this difference. As a result of the volume limitations of IllustrisTNG, we explore the cross-correlation function at significantly smaller scales than what is

typically used in void studies ($r \sim 100 \text{ Mpc } h^{-1}$). This creates some difficulty in drawing robust conclusions from our analysis, and we defer more extensive comparisons at larger separation to future work.

4 DISCUSSION

Our results, along with other work in the literature, demonstrate that physical models of galaxy formation naturally predict a modulation of the galaxy HOD with parameters other than halo mass, which has a significant impact on clustering. However, this work taken together demonstrates that the degree of galaxy assembly bias, and its physical origin, is somewhat model dependent. One advantage of SAMs is that the models are relatively simple, so it is a bit easier to pull them apart and determine the physical origin of observed effects, unlike hydrodynamic simulations which are more difficult to interpret. For example, the SC-SAM does not currently include any explicit environment-dependent physics in the models, so any observed dependence on environment must arise indirectly through the dependence of halo merger history on environment. In contrast, the galaxies in TNG experience different tidal forces in different environments, and may also be affected by ram pressure stripping, harassment, or merging, all of which depend strongly on environment in a direct way. Indeed, we showed that the gas fraction in $10^{13} M_{\odot}$ haloes in TNG shows significant dependence on environment. It is interesting therefore that for central galaxies, the dependence of the HOD on environment in the SAMs is of a similar magnitude to that seen in TNG. However, the dependence of the satellite HOD on environment is much stronger in TNG than in the SAMs, which is not surprising, as all of the environmental effects listed above are expected to work mainly on satellites. On the other hand, the SAMs show a much stronger dependence of the HOD on halo intrinsic properties, such as concentration and assembly time, than TNG. This may be because some of the SAM physics prescriptions more explicitly contain these quantities or closely related ones. For example, the mass outflow rate of stellar driven winds in TNG is an explicit function of the kernel weighted DM velocity dispersion within the nearest 64 particles, while in the SAM, the corresponding quantity is parametrized in terms of the galaxy-scale disc velocity, estimated as the halo rotation velocity at a radius of two NFW scale radii (which is directly related to concentration). Another example is that in the SAM, the galaxy disc scale radius is modelled using a formula that contains a direct dependence on both concentration and spin, while numerical simulations generally do not show a strong correlation between galaxy size and halo spin (e.g. Jiang et al. 2019). Work in progress (Karmakar et al., in preparation) demonstrates that galaxy size and halo spin also show no significant correlation in the TNG simulations.

The work presented here therefore points to the interesting result that galaxy-scale properties may be imprinted on large-scale galaxy clustering, through their impact on baryonic physics and the related influence on the HOD. We plan to use SAMs to further explore this by varying the parametrization of key physical processes such as stellar driven winds. Considered from another angle, it is clear that it is imperative to gain a better understanding of how to model the sub-grid physical processes in cosmological simulations, in a manner that faithfully captures the correct dependencies. Efforts such as SMAUG,² which aims to build up physical prescriptions for sub-grid models based on small-scale simulations that resolve

²Simulating Multiscale Astrophysics to Understand Galaxies; <https://www.simonsfoundation.org/flatiron/center-for-computational-astronomy/galaxy-formation/smaug/>

the relevant physical processes, should help enable progress on this problem.

An additional difference between the two models is that the SC-SAM assumes that the hot gas halo is stripped instantaneously when a halo enters the virial radius and becomes a sub-halo. As a result, satellite galaxies have no source of gas accretion, and quickly consume their gas by turning it into stars or expel it via stellar driven winds. A more realistic picture is probably that the hot gas haloes are gradually stripped over a finite time-scale, while cold gas may also to some extent be stripped out via ram pressure and tidal forces. A further uncertainty regarding satellite evolution is that they may be tidally destroyed via interaction with (especially) the massive central galaxy as well as perhaps other satellites. In this study, we attempted to circumvent these issues by making the stellar-mass selections separately for centrals and satellites, and by placing the satellites in the SAM at the same locations as the satellites in TNG. However, in the future, the SC-SAM should be updated to accurately account for these processes, using the results from numerical simulations as a guide. This has already been done for several SAMs such as GALFORM (Font et al. 2008), SAG (Tecce et al. 2010; Cora et al. 2018), DARK SAGE (Stevens, Croton & Mutch 2016), which have implemented gradual gas stripping for satellites, and L-GALAXIES (Ayromlou et al. 2019, 2021b), which has stripping for both centrals and satellites.

Another potential complication arises from the use of different halo catalogues – the default halo finder for TNG is FoF based, while the SAMs are built on halo merger trees based on the ROCKSTAR spherical overdensity based halo finder, although previous analysis in the literature (Gómez et al. 2021) suggests that the overall effect on galaxy properties when adopting different halo finders should be small as long as the merger trees are processed to behave in a physically robust manner. We show that different halo finders can give rise to moderate differences in the assembly bias signatures for DM haloes, complicating the interpretation of results in the literature which may have used different halo finders among other differences. However, we made use of bijective matches between the haloes generated by both methods, which allowed us to study the effect of this choice on our results. As shown in Fig. A2, this does not appear to affect our conclusions significantly.

In order to make an unbiased inference of cosmological parameters, it is essential that assembly bias is properly accounted for in the construction of mock catalogues. For example, Leauthaud et al. (2017) showed that mock catalogues constructed with standard HOD techniques to match the clustering of the CMASS sample predict a galaxy–galaxy lensing signal that is 20–40 per cent larger than observed, and suggested that this could be due to shortcomings in our understanding of galaxy formation, or to new physics. Recent work has shown that modifying the standard HOD formalism to include assembly bias does reduce the ‘lensing is low’ discrepancy (e.g. Zu 2020; Lange et al. 2021; Yuan et al. 2021). The ability of relatively computationally efficient physics-based models, such as SAMs, to qualitatively reproduce the assembly bias signal seen in more sophisticated numerical hydrodynamic simulations is very promising for developing techniques to create large volume mock galaxy catalogues that are consistent with both lensing and galaxy clustering as well as other cosmological probes.

In order to grasp the origin of the differences in clustering predictions between SAMs and hydro simulations, it is fundamental to understand how these two techniques populate dark matter haloes with different galaxy populations. As an example, the SC-SAM and TNG model baryonic processes and gravitational processes such as tidal stripping and destruction adopting completely distinct

frameworks. We speculate that differences in the predicted rate at which gas can cool and accrete into haloes, the assumed scalings for stellar feedback and AGN feedback, and the calibration choices, all contribute. We are pursuing a deeper understanding of the relative importance of these factors in ongoing work.

In future work, we plan to retool the SC-SAMs to deliberately mimic the sub-grid physics recipes in the TNG simulations, including potentially tuning the calibration parameters to match TNG instead of observations. It is clear from the work presented here that attention should be paid to environment-dependent physical processes and especially their impact on satellite galaxies, such as tidal and ram pressure stripping, which will probably need to be included in the SAM. Other interesting questions to be explored include the dependence of assembly bias on redshift and on different tracers used to select galaxies, such as SFR or emission line based selection. Clearly our ultimate goal is to develop methods to produce mock catalogues with multiple observable tracers that reproduce the observed clustering properties over a broad range of redshift and for different tracers and galaxy selection criteria.

Finally, we recommend including more large-scale statistical probes in the validation and perhaps calibration of SAMs and simulations. We showed that although the SC-SAM and TNG produce very similar predictions for two-point statistics on large scales, they show interesting differences in higher order statistics such as the third moment of the smoothed density and void statistics. As we gain the ability to populate larger volumes with galaxies, observational measurements of these higher order clustering statistics will be invaluable for more stringent validation of the next generation of galaxy models.

4.1 Our results in the context of other studies in the literature

Some of the earliest studies on assembly bias in the literature have looked at the dependence of halo clustering on parameters beyond halo mass in dark-matter-only simulations (e.g. Gao et al. 2005; Gao & White 2007). Gao et al. (2005) found that haloes that assemble at early times are substantially more clustered than their late-time counterparts. In Gao & White (2007), the study was extended to include a larger number of parameters at different redshifts such as halo concentration and spin. These early papers were among the first to suggest that halo occupation depends on the large-scale environment of the halo and that assembly bias effects may lead to differences at the 10 per cent level. This study agrees qualitatively with the impact of assembly history on the halo clustering and confirms that in both TNG and SC-SAM, not accounting for secondary parameter dependencies leads to a discrepancy of order 10 per cent.

More recent papers have explored the dependence of halo occupancy on additional halo parameters in different physics-based models of galaxy formation. For example, Zehavi et al. (2018) investigate the effect of large-scale environment and halo formation time on galaxy samples obtained from a semi-analytic galaxy model. They find that early-forming haloes are more likely to host central galaxies at lower halo mass, while the opposite is true for the satellites. Artale et al. (2018), on the other hand, perform a similar analysis, but this time employing the two hydrodynamical simulations, EAGLE and Illustris, reporting that haloes in the densest environments are more likely to host a central galaxy than those in the least dense environments. This relationship is found to be even stronger for early-forming haloes that have had more time to assemble. This study similarly finds that halo occupation exhibits noticeable dependence on both environment and formation time for TNG and SC-SAM, with TNG having more sensitivity to environment and SC-SAM to

formation time. Ayromlou et al. (2021a) reached a similar conclusion, finding that the baryon fractions of L-GALAXIES and TNG differ noticeably, most likely due to differences in the cooling and feedback models. In addition, they report that environmental effects seen in TNG are not captured accurately in SAMs particularly around clusters and groups.

Formation time and concentration, although extensively studied, have been shown, however, to be insufficient for reproducing the full galaxy assembly bias signal (e.g. Croton et al. 2007; Hadzhiyska et al. 2020). Several authors have also shown that augmenting galaxy population models with environmental information for each halo can help us to fully reproduce the galaxy assembly bias signal (e.g. Contreras et al. 2021; Xu, Zehavi & Contreras 2021; Hadzhiyska et al. 2021a). Bose et al. (2019) offer a careful consideration of the secondary parameters affecting halo occupation in TNG and their complex interplay, whereas Mao et al. (2018) consider the relationship between halo parameters in the context of the clustering of cluster-size haloes.

A work similar in spirit to ours was presented in Contreras et al. (2021), where the authors examined galaxy assembly bias on both a stellar-mass and a star-formation selected sample using TNG, a semi-analytic model, and a SHAM. While here we only focus on mass-selected populations, we put a significant effort into understanding the assembly bias properties of each model as well as the response of the galaxy clustering and halo occupation to a large variety of parameters. An additional difference is that our study includes statistics beyond the two-point correlation function such as moments of the density field and void statistics. An interesting observation is that SC-SAM and TNG appear to exhibit much more similar large-scale clustering behaviour and galaxy assembly bias than the SAM and hydro samples in Contreras et al. (2021). This may be attributed to differences in the SAMs employed in both studies and also the sample selection choices.

A thorough study of the galaxy clustering and galaxy-galaxy lensing signal at smaller scales ($r \lesssim 2 \text{ Mpc}/h$) was conducted by Renneby et al. (2020), in which the authors compare IllustrisTNG and the L-GALAXIES SAMs with observational data on two-point clustering statistics. They find that TNG300 makes predictions for the lensing of stellar-mass-selected objects, locally bright galaxies (LBGs), and red galaxies that are in excellent agreement with observations, but struggles with the highest mass bins for the blue population. Contrary to the findings in our work for which we find relatively good agreement between SC-SAM and TNG, they report that the unmodified L-GALAXIES SAMs perform significantly worse than TNG. However, minimizing the residual differences between the SAM and TNG galaxy properties while varying SAM parameters, the authors find more optimal parameters for the dynamical friction and AGN feedback that provide a comparable fit to the data. Furthermore, they note that low convergence and resolution can significantly impact the clustering predictions in hydro simulations and suggest implementing resolution-independent recipes for star formation. Significant challenge present in TNG and the best-fitting SAMs is predicting $\Delta \Sigma(r)$ for red galaxies in SDSS data with stellar masses $10^{9.4} - 10^{11} M_{\odot}$.

5 SUMMARY AND CONCLUSIONS

In the last decade, cosmologists have been faced with the task of populating increasingly larger N -body simulations with galaxies to interpret the wealth of available and upcoming galaxy observations. There is established consensus that the most accurate way to do this involves ab-initio prescriptions, which meticulously trace and evolve

the various components of the Universe that contribute to galaxy formation according to the physical laws that govern them. Existing empirical recipes, such as the HOD and SHAM, make numerous simplifying assumptions; they not only lack physical motivation, but also are inaccurate at the 10 percent level for certain tracers of the large-scale structure. On the other hand, SAMs capture the essential physics of galaxy formation and output multiple galaxy properties, such as star formation rate and mass contained in stars, cold interstellar medium, and warm circumgalactic medium. They are thus a promising venue for modelling the connection between galaxies and the underlying matter distribution.

This work uses the latest version of the well-established Santa Cruz SAM, as run in merger trees extracted from the dark matter only version of the IllustrisTNG simulations (Gabielpillai et al., in preparation), and compares various large-scale properties of galaxies in the full-physics TNG simulations to those in the Santa-Cruz SAM (SC-SAM) at $z = 0$ for a stellar-mass-selected sample. We pointed out differences and similarities between the two models, and compared their assembly bias signatures.

The overall stellar-mass versus halo mass (SMHM) relation and HOD for the SAMs and TNG are similar, but show some discrepancies. TNG tends to produce more massive galaxies in haloes with $M_{\text{halo}} \gtrsim 10^{13} M_{\odot} h^{-1}$ than the SC-SAM. This is due to differences in the AGN feedback physics recipes, and is partially a choice of calibration – it is straightforward to reproduce more TNG-like results in the SC-SAM by reducing the strength of ‘radio mode’ AGN feedback. Something similar has already been done in Renneby et al. (2020), where the authors modify the ‘radio mode’ AGN feedback in order to obtain better agreement with TNG for galaxy clustering and galaxy–galaxy lensing. The HOD turns over more gradually in the SAM and has a longer ‘tail’ of galaxies populating lower mass haloes. This is largely due to the sharper peak in the stellar-mass versus halo mass relation and the larger scatter in this relation for halo masses around the peak ($M_{\text{halo}} \gtrsim 10^{12} M_{\odot}$) in the SAM relative to TNG. Interestingly, SAMs seem to generically predict a larger scatter in the SMHM relation than numerical simulations (Wechsler & Tinker 2018), for reasons that are not well-understood. The scatter in the SMHM relation at these halo masses is also not well constrained from observations, though results presented in Cao et al. (2020) suggest that the scatter increases towards lower halo masses in a manner more similar to the SAM predictions.

A major finding is the excellent agreement between the SC-SAM and TNG for the two-point correlation function on scales beyond $r > 2 \text{ Mpc } h^{-1}$ (Fig. 4), galaxy assembly bias signature (Fig. 5) and qualitative response to the various halo parameters examined, seen in both the HOD as well as the galaxy assembly bias (Figs 3 and 6). This is despite the fact that the SAM was independently developed, with physical prescriptions that differ from those of TNG in many respects, and was calibrated to a different set of observations. Such results are encouraging; they argue that the SC-SAM model to first order correctly captures many of the dominant features of the galaxy-halo connection produced by a much more computationally expensive method. Our results constitute direct evidence that SAMs are capable of capturing the dependence of clustering on halo parameters other than mass, thereby providing a physically motivated improvement over standard HOD models.

However, there are some interesting and puzzling differences in the relationships between halo occupation and clustering with secondary parameters between the SAM and TNG. The SC-SAM and TNG central galaxy HODs show a similar degree of sensitivity to environment, but the SAMs show much greater sensitivity to other parameters such as concentration and halo formation time

than does TNG. The clustering in both the SAM and TNG shows a significant sensitivity to environment when HODs are redistributed among haloes that are ranked according to this parameter. TNG shows a weak sensitivity to internal parameters such as concentration and formation time in the ranked shuffling test, while clustering in the SAMs shows greater sensitivity to these parameters. As noted above, we speculate that this may be due to differences in the details of the quantities that are used to parametrize feedback recipes in the two models (e.g. halo scale versus galaxy scale velocity).

In order to further explore the physical reasons for the differences seen between the models, we investigated the gas fractions on two different scales for haloes of the same mass $M_{\text{halo}} \sim 10^{13} M_{\odot} h^{-1}$ but in different environments in TNG, and compared this with the same quantity measured within the halo virial radius for SAMs. We found a significantly greater variation in gas fraction between low and high density environments in TNG than in the SAM, which showed virtually no dependence of this quantity on environment.

Emulating state-of-the-art hydro simulations, such as IllustrisTNG, with ‘cheaper’ methods that incorporate the basic galaxy formation model, such as SAMs, would revolutionize the development of pipelines for cosmological surveys and change the face of precision cosmology. In the next few years, the extent to which SAMs can reproduce hydro simulations will be put to the test, and exciting advancements in cosmology are bound to come about!

ACKNOWLEDGEMENTS

We would like to thank Sihan Yuan and the anonymous referee for illuminating discussions. BH and DJE were supported by DOE-SC0013718 and NASA ROSES grant 12-EUCLID12-0004. RSS gratefully acknowledges support from the Simons Foundation. SB acknowledges support from Harvard University through the ITC Fellowship. DJE is further supported in part as a Simons Foundation investigator. Some of the simulations used in this work were produced using computing facilities provided by the Flatiron Institute.

DATA AVAILABILITY

The IllustrisTNG data is publicly available at www.tng-project.org. The SAM outputs will be made available upon reasonable request to the authors. The scripts used in this project are available at https://github.com/boryanah/SAM_TNG_clustering.

REFERENCES

Abbas U., Sheth R. K., 2007, *MNRAS*, 378, 641
 Alam S., Zu Y., Peacock J. A., Mandelbaum R., 2019, *MNRAS*, 483, 4501
 Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, *MNRAS*, 480, 3978
 Ayromlou M., Kauffmann G., Yates R. M., Nelson D., White S. D. M., 2021b, *MNRAS*, 505, 492
 Ayromlou M., Nelson D., Yates R. M., Kauffmann G., Renneby M., White S. D. M., 2021a, *MNRAS*, 502, 1051
 Ayromlou M., Nelson D., Yates R. M., Kauffmann G., White S. D. M., 2019, *MNRAS*, 487, 4313
 Barnes J., Efstathiou G., 1987, *ApJ*, 319, 575
 Baugh C. M., 2008, *Phil. Trans. R. Soc. A*, 366, 4381
 Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, *ApJ*, 762, 109
 Behroozi P. S., Wechsler R. H., Conroy C., 2013b, *ApJ*, 770, 57
 Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, 488, 3143
 Beltz-Mohrmann G. D., Berlind A. A., Szeuwiw A. O., 2020, *MNRAS*, 491, 5771
 Benson A. J., 2017, *MNRAS*, 471, 2871

Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, *MNRAS*, 311, 793
 Bernardeau F., 1994, *A&A*, 291, 697
 Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, *MNRAS*, 376, 215
 Bose S., Eisenstein D. J., Hernquist L., Pillepich A., Nelson D., Marinacci F., Springel V., Vogelsberger M., 2019, *MNRAS*, 490, 5693
 Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
 Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001a, *MNRAS*, 321, 559
 Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001b, *ApJ*, 555, 240
 Cao J.-z., Tinker J. L., Mao Y.-Y., Wechsler R. H., 2020, *MNRAS*, 498, 5080
 Chaves-Montero J., Angulo R. E., Schaye J., Schaller M., Crain R. A., Furlong M., Theuns T., 2016, *MNRAS*, 460, 3100
 Cochrane R. K., Best P. N., 2018, *MNRAS*, 480, 864
 Contreras S., Angulo R. E., Zennaro M., 2021, *MNRAS*, 504, 5205
 Contreras S., Zehavi I., Padilla N., Baugh C. M., Jiménez E., Lacerna I., 2019, *MNRAS*, 484, 1133
 Cora S. A. et al., 2018, *MNRAS*, 479, 2
 Croton D. J. et al., 2016, *ApJS*, 222, 22
 Croton D. J., Gao L., White S. D. M., 2007, *MNRAS*, 374, 1303
 Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
 Desjacques V., Jeong D., Schmidt F., 2018, *Phys. Rep.*, 733, 1
 Diemer B., Kravtsov A. V., 2015, *ApJ*, 799, 108
 Dutton A. A., Macciò A. V., 2014, *MNRAS*, 441, 3359
 Favole G., Rodríguez-Torres S. A., Comparat J., Prada F., Guo H., Klypin A., Montero-Dorta A. D., 2017, *MNRAS*, 472, 550
 Font A. S. et al., 2008, *MNRAS*, 389, 1619
 Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
 Gao L., White S. D. M., 2007, *MNRAS*, 377, L5
 Gaztanaga E., 1994, *MNRAS*, 268, 913
 Genel S. et al., 2014, *MNRAS*, 445, 175
 Gómez J. S., Padilla N. D., Helly J. C., Lacey C. G., Baugh C. M., Lagos C. D. P., 2021, preprint ([arXiv:2106.12664](https://arxiv.org/abs/2106.12664))
 Gonzalez-Perez V. et al., 2018, *MNRAS*, 474, 4024
 Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, *MNRAS*, 439, 264
 Gregory S. A., Thompson L. A., 1978, *ApJ*, 222, 784
 Guo H. et al., 2016a, *MNRAS*, 459, 3040
 Guo Q. et al., 2016b, *MNRAS*, 461, 3457
 Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, *MNRAS*, 428, 1351
 Hadzhiyska B., Bose S., Eisenstein D., Hernquist L., 2021a, *MNRAS*, 501, 1603
 Hadzhiyska B., Bose S., Eisenstein D., Hernquist L., Spergel D. N., 2020, *MNRAS*, 493, 5506
 Hadzhiyska B., Tacchella S., Bose S., Eisenstein D. J., 2021b, *MNRAS*, 502, 3599
 Hayashi E., White S. D. M., 2008, *MNRAS*, 388, 2
 Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, *MNRAS*, 451, 2663
 Hopkins P. F. et al., 2018, *MNRAS*, 480, 800
 Jiang F. et al., 2019, *MNRAS*, 488, 4801
 Johnson J. W., Maller A. H., Berlind A. A., Sinha M., Holley-Bockelmann J. K., 2019, *MNRAS*, 486, 1156
 Klypin A. A., Trujillo-Gomez S., Primack J., 2011, *ApJ*, 740, 102
 Knebe A. et al., 2011, *MNRAS*, 415, 2293
 Kreisch C. D., Pisani A., Carbone C., Liu J., Hawken A. J., Massara E., Spergel D. N., Wandelt B. D., 2019, *MNRAS*, 488, 4413
 Lange J. U., Leauthaud A., Singh S., Guo H., Zhou R., Smith T. L., Cyr-Racine F.-Y., 2021, *MNRAS*, 502, 2074
 Leauthaud A. et al., 2017, *MNRAS*, 467, 3024
 Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, *MNRAS*, 460, 1214
 Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, *MNRAS*, 441, 378
 Mao Y.-Y., Zentner A. R., Wechsler R. H., 2018, *MNRAS*, 474, 5143
 Marinacci F. et al., 2018, *MNRAS*, 480, 5113

- Masaki S., Hikage C., Takada M., Spergel D. N., Sugiyama N., 2013, *MNRAS*, 433, 3506
- More S., Kravtsov A. V., Dalal N., Gottlöber S., 2011, *ApJS*, 195, 4
- Moster B. P., Naab T., White S. D. M., 2013, *MNRAS*, 428, 3121
- Moster B. P., Naab T., White S. D. M., 2018, *MNRAS*, 477, 1822
- Naiman J. P. et al., 2018, *MNRAS*, 477, 1206
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Nelson D. et al., 2018, *MNRAS*, 475, 624
- Nelson D. et al., 2019, *MNRAS*, 490, 3234
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Oguri M., Takada M., 2011, *Phys. Rev. D*, 83, 023008
- Oñorbe J., Garrison-Kimmel S., Maller A. H., Bullock J. S., Rocha M., Hahn O., 2014, *MNRAS*, 437, 1894
- Pandya V. et al., 2020, *ApJ*, 905, 4
- Paranjape A., Hahn O., Sheth R. K., 2018, *MNRAS*, 476, 5442
- Paranjape A., Kovač K., Hartley W. G., Pahwa I., 2015, *MNRAS*, 454, 3030
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peebles P. J. E., 1969, *ApJ*, 155, 393
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton
- Pillepich A. et al., 2018a, *MNRAS*, 473, 4077
- Pillepich A. et al., 2018b, *MNRAS*, 475, 648
- Pillepich A. et al., 2019, *MNRAS*, 490, 3196
- Popping G. et al., 2019, *ApJ*, 882, 137
- Pujol A., Hoffmann K., Jiménez N., Gaztañaga E., 2017, *A&A*, 598, A103
- Renneby M., Henriques B. M. B., Hilbert S., Nelson D., Vogelsberger M., Angulo R. E., Springel V., Hernquist L., 2020, *MNRAS*, 498, 5804
- Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, *MNRAS*, 462, 893
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, *ApJ*, 546, 20
- Seljak U. et al., 2005, *Phys. Rev. D*, 71, 043511
- Shi J., Sheth R. K., 2018, *MNRAS*, 473, 2486
- Sinha M., Garrison L. H., 2020, *MNRAS*, 491, 3022
- Somerville R. S. et al., 2021, *MNRAS*, 502, 4858
- Somerville R. S., Davé R., 2015, *ARA&A*, 53, 51
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, *MNRAS*, 391, 481
- Somerville R. S., Popping G., Trager S. C., 2015, *MNRAS*, 453, 4337
- Somerville R. S., Primack J. R., 1999, *MNRAS*, 310, 1087
- Spearman C., 1904, *Am. J. Psychol.*, 15, 88
- Springel V. et al., 2018, *MNRAS*, 475, 676
- Springel V., 2010, *MNRAS*, 401, 791
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 289
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Stevens A. R. H., Croton D. J., Mutch S. J., 2016, *MNRAS*, 461, 859
- Sunayama T. et al., 2020, *MNRAS*, 496, 4468
- Sutherland R. S., Dopita M. A., 1993, *ApJS*, 88, 253
- Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, *ApJ*, 868, 92
- Tacchella S., Trenti M., Carollo C. M., 2013, *ApJ*, 768, L37
- Tecce T. E., Cora S. A., Tissera P. B., Abadi M. G., Lagos C. D. P., 2010, *MNRAS*, 408, 2008
- Vogelsberger M. et al., 2014a, *MNRAS*, 444, 1518
- Vogelsberger M. et al., 2014b, *Nature*, 509, 177
- Wang E., Wang H., Mo H., van den Bosch F. C., Lim S. H., Wang L., Yang X., Chen S., 2018, *ApJ*, 864, 51
- Wechsler R. H., Bullock J. S., Primack J. R., Kravtsov A. V., Dekel A., 2002, *ApJ*, 568, 52
- Wechsler R. H., Tinker J. L., 2018, *ARA&A*, 56, 435
- Weinberger R. et al., 2017, *MNRAS*, 465, 3291
- Weinberger R., Springel V., Pakmor R., 2020, *ApJS*, 248, 32
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Xu X., Zehavi I., Contreras S., 2021, *MNRAS*, 502, 3242
- Xu X., Zheng Z., 2020, *MNRAS*, 492, 2739
- Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., Chu Y., 2004, *MNRAS*, 350, 1153
- Yuan S., Hadzhiyska B., Bose S., Eisenstein D. J., Guo H., 2021, *MNRAS*, 502, 3582
- Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019, *MNRAS*, 483, 2983
- Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, *ApJ*, 853, 84
- Zheng Z. et al., 2005, *ApJ*, 633, 791
- Zu Y., 2020, preprint ([arXiv:2010.01143](https://arxiv.org/abs/2010.01143))

APPENDIX A: SENSITIVITY TO HALO FINDERS

In this section, we discuss the effects of utilizing two distinct halo-finding catalogues in our comparison between TNG and SC-SAM. As discussed in Section 2.1, SC-SAM is constructed on top of a merger tree catalogue based on haloes identified using the ROCKSTAR halo finder adopting the default ROCKSTAR parameters.³ TNG halo and subhalo properties, on the other hand, are reported in reference to catalogues obtained via the FoF and SUBFIND algorithms. It is known that different halo finders can yield different results for halo property distributions (Klypin, Trujillo-Gomez & Primack 2011; Knebe et al. 2011).

As an initial comparison, we show the halo mass function of the ROCKSTAR and FoF/SUBFIND catalogues in Fig. A1. As halo mass proxy in TNG, we adopt ‘Group_M_TopHat200’, as it uses the same virial mass definition (Bryan & Norman 1998) as the default choice in ROCKSTAR. Despite that, we see that the discrepancy in the number of haloes between the two catalogues is ~ 15 per cent for halo masses between 10^{10} and $10^{13} M_{\odot} h^{-1}$. At higher halo masses, the difference is smaller, and the ratio approaches unity for the largest haloes. The most likely explanation for the number of ROCKSTAR haloes being larger is that the ROCKSTAR finder uses 6D information to identify haloes and thus deblends many of the large percolating FoF groups. This is particularly noticeable at higher densities, where the FoF finder tends to string together distinct dark-matter structures into a single FoF object particularly when they are orbiting on the outskirts of large dark-matter clumps.

In an effort to correct for the fact that TNG reports FoF halo properties, in Figs 3 and 6, we make use of the bijective matches provided between the SC-SAM and TNG catalogues. In addition, here we explore the effect of associating the TNG galaxies with ROCKSTAR haloes. We do this via the following procedure: for each TNG subhalo, we find the most massive ROCKSTAR halo that encompasses it within its virial radius and assign it to that halo. For the stellar-mass-selected sample at $z = 0$, we find matches for 99.7 per cent of the TNG galaxies. We see that the HOD of the TNG galaxies remains almost unchanged. This implies that the root cause of the differences between TNG and the SAM is likely unrelated to the halo finding, but is rather caused by a different implementation of the galaxy formation processes and the resulting galaxy–halo connection.

³For more information, see <https://github.com/yt-project/rockstar/blob/master/README.md>.

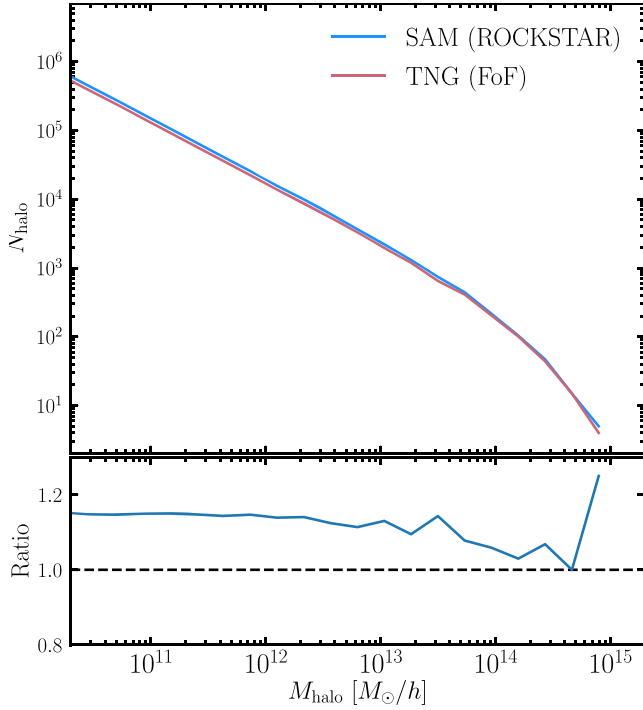


Figure A1. Halo mass function (HMF) of haloes identified with ROCKSTAR (blue) and with FoF/SUBFIND (red). We adopt the default virial mass definition from ROCKSTAR, while for FoF/subfind, we adopt ‘Group_M_TopHat200’, which uses the same virial mass definition (Bryan & Norman 1998). The lower panel shows the ratio between the two and indicates that ROCKSTAR identifies 15 per cent more haloes on the mass scale between 10^{10} and $10^{13} M_{\odot} h^{-1}$.

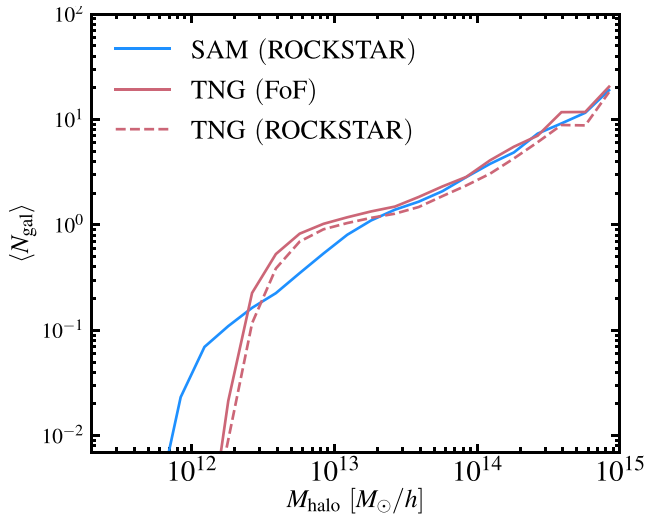


Figure A2. Halo occupation distribution for a stellar-mass-selected sample with a galaxy number density of $n_{\text{gal}} = 0.001 [\text{Mpc } h^{-1}]^{-3}$ (corresponding to 12 000 galaxies) at $z = 0$ at $n_{\text{gal}} = 0.001 [\text{Mpc } h^{-1}]^{-3}$. In blue, we show the HOD of the SAM galaxy sample, which is constructed from ROCKSTAR halo catalogues, while in red, we show the result for TNG. The default halo finder used to compute the halo statistics in TNG is FoF/SUBFIND. In dashed red, we show what the HOD would look like if we match the TNG galaxies to ROCKSTAR haloes and use the ROCKSTAR based mass instead. We see that the HOD of the TNG galaxies changes only marginally.

APPENDIX B: HALO ASSEMBLY BIAS AND PARAMETER CORRELATIONS

In this section, we study the strength of the correlation between various halo parameters that are relevant in the study of galaxy assembly bias. We also visualize the halo assembly bias signal resulting from halo concentration and environment. These relations have been extensively studied and reported on in the literature, so we present them here largely as a baseline.

In Fig. B1, we show 70 per cent and 95 per cent 2D contours, illustrating the scatter plots of a few pairs of halo parameters, weighted using kernel density estimation (KDE). More information about the parameters, including their definitions (environment, concentration, formation epoch, spin, velocity anisotropy, maximum circular velocity, FoF to virial mass ratio), can be found in Section 2.4. The haloes are split into three mass ranges: $\log M_{\text{halo}} = 12.0$ – 12.5 , 12.5 – 13.0 , and 13.0 – 13.5 , in units of $M_{\odot} h^{-1}$. We also quote the Spearman’s rank coefficient, r_s , for each mass bin (Spearman 1904). To get a sense of the correlation strength, we remind the reader that the ranges $0 \leq r_s < 0.2$, $0.2 \leq r_s < 0.4$, $0.4 \leq r_s < 0.6$, and $0.6 \leq r_s < 0.8$ correspond to very weak, weak, moderate, and strong correlation. The strongest correlation is between the formation time and concentration, where as expected early-forming haloes are much more concentrated. The correlation between spin and concentration is also relatively stronger compared with that between other pairs and confirms the findings of other authors (e.g. Bose et al. 2019). Interestingly, environment is very weakly correlated with concentration and spin, so including it as an assembly bias parameter would not introduce a strong degeneracy in the halo model. Environment demonstrates the strongest amount of correlation with velocity anisotropy for low-mass haloes and the ratio of the FoF to virial mass for high-mass haloes. A likely reason for this finding is that high-environment haloes live in dense regions with a large number of mergers (hence lower velocity anisotropy), where the FoF group percolates over a large distance.

To estimate the halo assembly bias, we split the haloes into 6 mass bins: $\log M = 12.0 \pm 0.3$, 12.5 ± 0.3 , 13.0 ± 0.3 , 13.5 ± 0.3 , 14.0 ± 0.4 , 14.5 ± 0.5 . We define halo assembly bias as the ratio between the correlation function of the top/bottom 25 per cent haloes ordered by the parameter of choice, divided by the correlation function of the median population (50 per cent \pm 12.5 per cent). The ROCKSTAR haloes are abundance matched to the TNG-selected haloes for each mass bin to make the comparison more direct. The resulting comparison is shown in Fig. B2. For small haloes with $M_{\text{halo}} \sim 10^{12} M_{\odot} h^{-1}$, we see that high concentration results in stronger clustering, whereas this relation is reversed for the most massive haloes as expected based on previous studies in the literature. On the other hand, in the case of environment, across all mass bins haloes living in denser environments are more clustered than haloes living in less dense regions. As discussed in the main text, haloes in dense regions experience more mergers and can be part of extended cosmic web structures such as walls, filaments, and knots that have typical size of $\sim 10 \text{ Mpc } h^{-1}$. Comparing TNG with SAM, we see that the signal is stronger for SAM, confirming the conjecture that ROCKSTAR tends to deblend haloes more readily and identify dark matter structure on the outskirts of large haloes. It is also noteworthy that the halo assembly bias response to environment is significantly stronger than the response to concentration.

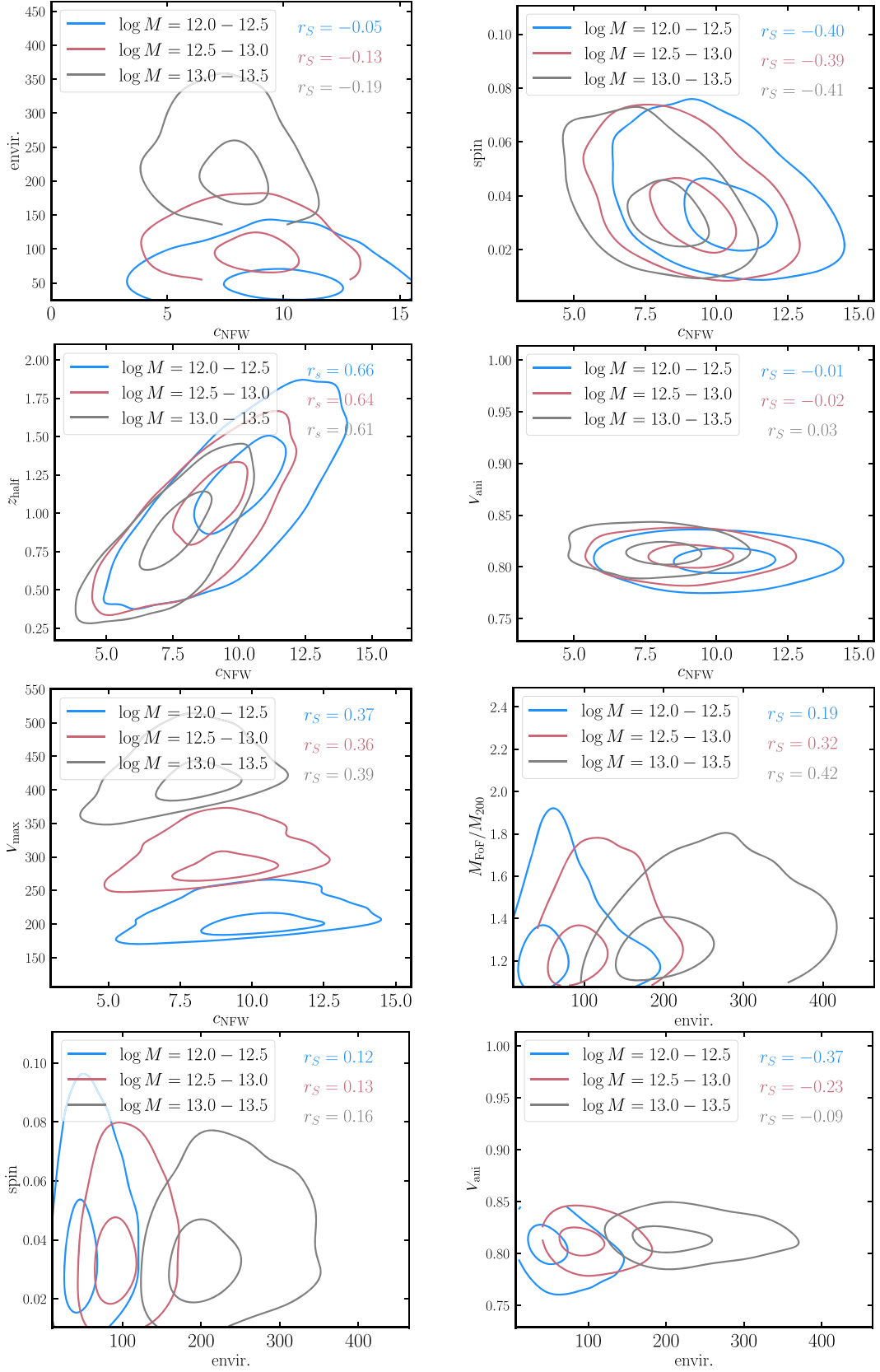


Figure B1. Correlations between the different halo parameters for 3 mass bins ($\log M_{\text{halo}} = 12.0 - 12.5$, $12.5 - 13.0$, and $13.0 - 13.5$) in units of $M_{\odot} h^{-1}$. Spearman's rank coefficients, r_S , are shown for each bin in colours corresponding to their respective mass bins. Higher r_S values correspond to stronger correlations. The density contours correspond to the 70th and 95th percentile of the KDE-weighted data, respectively. We use the same parameter definitions as listed in Section 2.4.

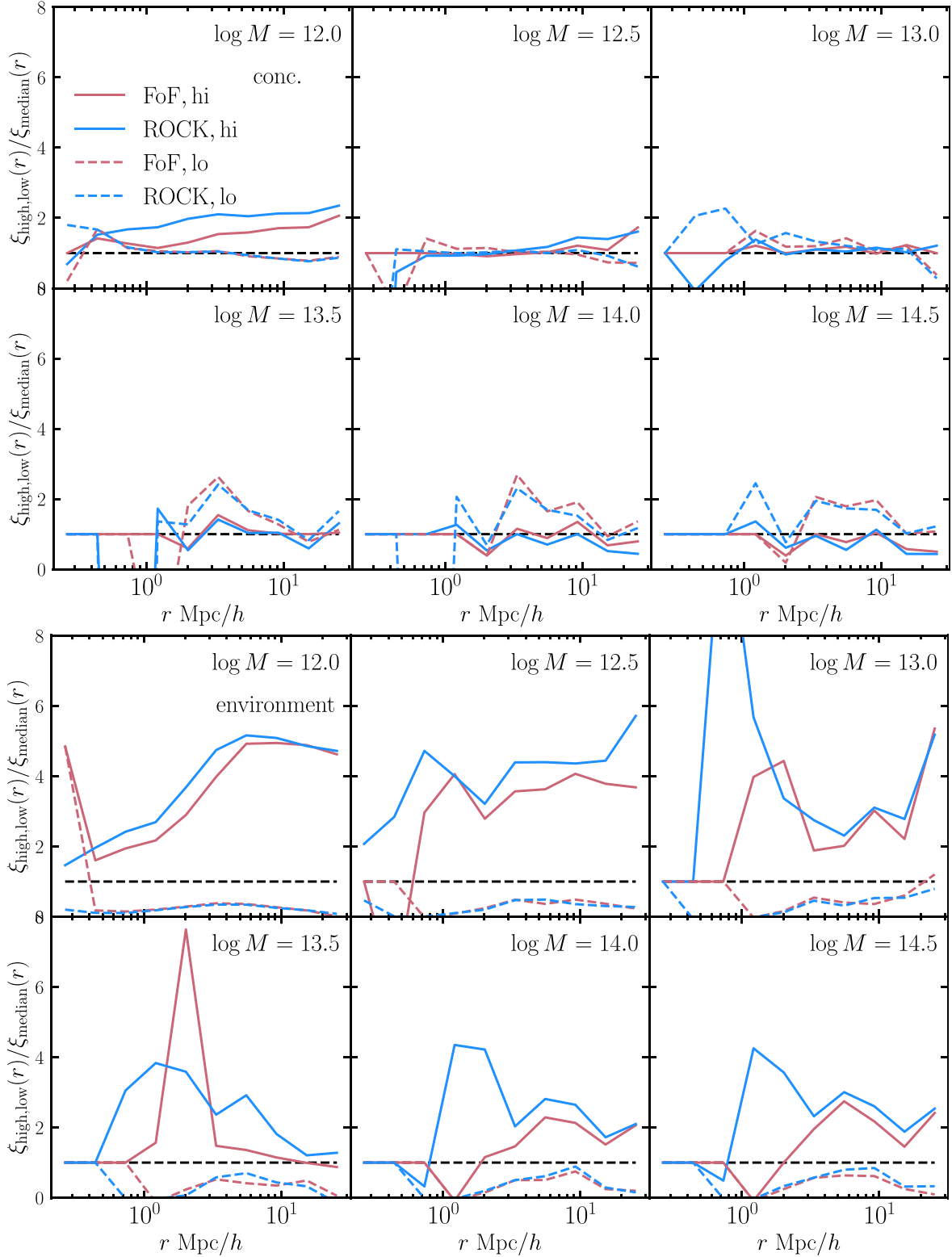


Figure B2. Halo assembly bias signature for the concentration (top) and environment (bottom) parameters for different halo masses ($\log M = 12.0, 12.5, 13.0, 13.5, 14.0, 14.5$). In blue, we show the halo assembly bias of the ROCKSTAR haloes (‘ROCK’), while in red, we show the result for FoF. The dotted lines correspond to the bottom 30 per cent of haloes (‘lo’), whereas the solid lines denote the top 30 per cent (‘hi’), when ordered by concentration/environment. The ROCKSTAR haloes are abundance matched to the TNG-selected haloes for each mass bin to make the comparison more direct. All masses are reported in units of $M_{\odot} h^{-1}$. The signal is much weaker in the concentration plot and for small halo masses, and we see that high concentration leads to positive assembly bias, whereas on the high mass end, the relation is inverted. On the other hand, in the case of environment, the bias signal is always stronger for haloes in high-density environments and weaker for haloes residing in low-density regions.